GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:10:03 ; Search time 5416.83 Seconds
(without alignments)
17274.699 Million cell updates/sec

Title: US-09-728-552A-3_COPY_1_2000
Perfect score: 2000
Sequence: 1 gaattctcctgcctcagcct.....tctgcaaaatagatgttgtg 2000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	415.8	20.8	432	9	AQ015472	AQ015472 CIT-HSP-2
c	2	413	20.6	8158	4	CR749268	CR749268 Homo sapi
	3	399.2	20.0	5937	4	HSM807294	BX647150 Homo sapi
	4	385	19.2	707	5	BU616791	BU616791 UI-H-FH1-
	5	373	18.6	700	5	BU622317	BU622317 UI-H-FH1-
	6	360.8	18.0	690	6	CA430123	CA430123 UI-H-FH1-
	7	360.6	18.0	9266	4	HSM806813	BX640738 Homo sapi
c	8	360.4	18.0	778	6	CB957195	CB957195 AGENCOURT
c	9	351.6	17.6	994	2	BG113167	BG113167 602283837
	10	341.2	17.1	814	1	AU120614	AU120614 AU120614
c	11	335.6	16.8	719	5	BX955959	BX955959 DKFZp781A
c	12	335.2	16.8	907	6	CF242986	CF242986 AGENCOURT
c	13	330.6	16.5	603	9	AQ320818	AQ320818 RPCI11-92
	14	330	16.5	687	9	BZ603609	BZ603609 WHAAV41TF
	15	330	16.5	718	5	BX495939	BX495939 DKFZp779B
c	16	329.8	16.5	721	7	CR538654	CR538654 DKFZp459M
	17	328.2	16.4	831	7	CK774525	CK774525 963293 MA
c	18	327.4	16.4	5336	4	CR859436	CR859436 Pongo pyg
c	19	326.6	16.3	929	2	BG476842	BG476842 602525095
c	20	325.6	16.3	755	9	CE040574	CE040574 tigr-gss-
c	21	325.6	16.3	815	9	AQ237450	AQ237450 RPCI11-61
	22	325.4	16.3	705	5	BU622127	BU622127 UI-H-FH1-
	23	325.2	16.3	844	9	AQ748474	AQ748474 HS_5540_A
c	24	325	16.2	849	9	CC478317	CC478317 CH240_304
c	25	324.2	16.2	1039	3	BQ067736	BQ067736 AGENCOURT
c	26	321.4	16.1	853	9	AQ750123	AQ750123 HS_5575_A
c	27	318	15.9	931	2	BG028427	BG028427 602294332
	28	316.4	15.8	693	6	CB853133	CB853133 UI-CF-FN0
c	29	315.8	15.8	703	6	CF126937	CF126937 UI-HF-ET0
c	30	314.8	15.7	822	10	CG734253	CG734253 RP11-57I1
	31	312.6	15.6	668	5	BU740723	BU740723 UI-E-EJ0-
c	32	311.8	15.6	774	5	BX094980	BX094980 BX094980
c	33	311.8	15.6	774	7	CR736370	CR736370 CR736370
	34	311.6	15.6	724	3	BM715720	BM715720 UI-E-EJ0-
	35	311.4	15.6	673	5	BX503658	BX503658 DKFZp686B
	36	311.2	15.6	671	10	CE455208	CE455208 tigr-gss-
	37	310.6	15.5	690	3	BM978279	BM978279 UI-CF-EC1
c	38	310.4	15.5	537	3	BP280947	BP280947 BP280947
c	39	306.6	15.3	1011	9	BZ771573	BZ771573 mct78a04.
	40	306	15.3	723	9	CC765346	CC765346 CH240_57C
c	41	305.6	15.3	746	9	CE027248	CE027248 tigr-gss-
c	42	305.4	15.3	674	2	BF981537	BF981537 602309408
	43	304.2	15.2	696	5	BU680222	BU680222 UI-CF-DU1
	44	304	15.2	849	8	DN100706	DN100706 1000083 M
c	45	303	15.2	620	6	CA308451	CA308451 UI-H-FT1-
c	46	301.2	15.1	945	10	CL423094	CL423094 RP11-733L
c	47	299	14.9	743	9	AZ519098	AZ519098 RPCI-11-7
	48	298.8	14.9	705	9	B93818	B93818 CIT-HSP-216
c	49	297.8	14.9	857	9	CC561017	CC561017 CH240_471
c	50	297.4	14.9	750	10	AG032008	AG032008 Pan trogl
c	51	295.8	14.8	682	1	AW813102	AW813102 RC3-ST018
c	52	295.4	14.8	646	2	BG542877	BG542877 602571052
	53	293.8	14.7	641	1	AV698979	AV698979 AV698979
	54	293.4	14.7	843	11	CR873464	CR873464 Sus scrof
c	55	293.2	14.7	657	10	AG034266	AG034266 Pan trogl
c	56	292.4	14.6	596	9	AQ379794	AQ379794 RPCI11-15
c	57	290.4	14.5	604	9	AQ345448	AQ345448 RPCI11-10
	58	289.8	14.5	556	9	BZ854056	BZ854056 CH240_281
	59	289.8	14.5	660	10	CE373153	CE373153 tigr-gss-
c	60	289.8	14.5	703	1	AW968835	AW968835 EST380911
c	61	289	14.4	693	9	AQ082065	AQ082065 RPCI11-54
	62	289	14.4	705	10	CE837321	CE837321 tigr-gss-
	63	289	14.4	853	9	CC562936	CC562936 CH240_474
	64	288.8	14.4	601	7	CO688813	CO688813 DG11-250m
c	65	286.8	14.3	800	2	BG771022	BG771022 602719272
c	66	285.8	14.3	588	9	AQ275021	AQ275021 RPCI-6-15
c	67	284.6	14.2	752	9	BZ774366	BZ774366 mcv93a08.
c	68	284.2	14.2	959	2	BF347229	BF347229 602021164
	69	283.4	14.2	698	1	AV726174	AV726174 AV726174
	70	282.4	14.1	617	1	AV691379	AV691379 AV691379
c	71	281.6	14.1	870	9	AQ896399	AQ896399 HS_3134_A
c	72	281.4	14.1	575	9	AQ371490	AQ371490 RPCI11-14
c	73	281	14.1	857	9	AQ750429	AQ750429 HS_5574_B
c	74	280.4	14.0	721	11	CR960685	CR960685 Homo sapi
c	75	279.8	14.0	566	9	AQ381021	AQ381021 RPCI11-16
c	76	278.6	13.9	632	1	AW813101	AW813101 RC3-ST018
c	77	278.4	13.9	657	9	AQ536886	AQ536886 RPCI-11-3
c	78	278.4	13.9	769	9	AQ751037	AQ751037 HS_5576_B
c	79	278.2	13.9	667	10	AG114200	AG114200 Pan trogl
c	80	278.2	13.9	723	10	AG094156	AG094156 Pan trogl
	81	278.2	13.9	781	9	CE248403	CE248403 tigr-gss-
	82	277.2	13.9	558	3	BM668477	BM668477 UI-E-CL1-
c	83	277.2	13.9	829	9	AQ748882	AQ748882 HS_5540_A
c	84	276.6	13.8	697	5	BX482990	BX482990 DKFZp686B
c	85	275.8	13.8	655	10	AG092291	AG092291 Pan trogl
c	86	275.4	13.8	709	10	AG132716	AG132716 Pan trogl
c	87	274.6	13.7	643	5	BX476722	BX476722 DKFZp686N
c	88	274.6	13.7	821	8	DN134483	DN134483 999699 MA
c	89	273.8	13.7	711	8	DN750100	DN750100 GL-Cf-689
c	90	272.6	13.6	722	9	CC060814	CC060814 MUGQ_CH25
	91	272.2	13.6	695	10	CE371129	CE371129 tigr-gss-
c	92	272	13.6	718	10	CL003669	CL003669 gbs1_VM5_
	93	271.8	13.6	707	9	BZ884524	BZ884524 CH240_232
c	94	271.6	13.6	613	10	CE665894	CE665894 tigr-gss-
c	95	271.4	13.6	740	11	CR972811	CR972811 Homo sapi

c	96	271	13.6	777	10	CE459394	CE459394 tigr-gss-
	97	271	13.6	826	7	CR763358	CR763358 DKFZp470I
c	98	270.2	13.5	845	9	AQ747199	AQ747199 HS_5539_A
	99	269.2	13.5	1058	3	BM476651	BM476651 AGENCOURT
	100	269	13.5	680	9	BZ772441	BZ772441 mcv36f09.

ALIGNMENTS

RESULT 1
AQ015472
LOCUS AQ015472 432 bp DNA linear GSS 09-JUN-1998
DEFINITION CIT-HSP-2303H17.TR CIT-HSP Homo sapiens genomic clone 2303H17, genomic survey sequence.
ACCESSION AQ015472
VERSION AQ015472.1 GI:3194208
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 432)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .432
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2303H17"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match 20.8%; Score 415.8; DB 9; Length 432;
Best Local Similarity 99.1%; Pred. No. 9.9e-56;
Matches 428; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy       1287 TGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAATATCA 1346
             | || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TACTCAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAATATCA 60

Qy       1347 GTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTT 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTT 120

Qy       1407 TATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTT 1465
             ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        121 TATGCACAGTAGAACTTCTTTGAAAATTGGAGTCAATCCTCTCAAACCCTGCTCTGCTTT 180

Qy       1466 AACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAG 1525
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAG 240

Qy       1526 TGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATA 1585
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATA 300

Qy       1586 AGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCT 1645
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCT 360

Qy       1646 TCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTC 1705
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        361 TCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGNTTCTACCACATCTGTGGTTC 420

Qy       1706 CTTCCTCCATTG 1717
             ||||||||||||
Db        421 CTTCCTCCATTG 432
```

RESULT 2
CR749268/c
LOCUS CR749268 8158 bp mRNA linear HTC 19-AUG-2004
DEFINITION Homo sapiens mRNA; cDNA DKFZp686B11107 (from clone DKFZp686B11107).
ACCESSION CR749268
VERSION CR749268.1 GI:51476267
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 8158)
AUTHORS Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
CONSRTM The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp686B11107) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686B11107
Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.
FEATURES Location/Qualifiers
source 1. .8158
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686B11107"
/tissue_type="fetal kidney"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host DH10B; sites SfiIA + SfiIB"
/dev_stage="fetal"
/note="zinc finger protein 621"
gene 1. .8158
/gene="DKFZp686B11107"
CDS 325. .1644
/gene="DKFZp686B11107"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH18124.1"
/db_xref="GI:51476268"
/translation="MLQTTWPQESVTFEDVAVYFTQNQWASLDPAQRALYGEVMLENY
ANVASLVAFPFPKPALISHLERGEAPWGPDPWDTEILRGISQGGESWIKNEGLVIKQE
ASEETELHRMPVGGLLRNVSQHFDFKRKALKQTFNLNPNLILRGGMKFYECKECGKIF
RYNSKLIRHQMSHTGEKPFKCKECGKAFKSSYDCIVHEKNHIGEGPYECKECGKGLSS
NTALTQHQRIHTGEKPYECKECGKAFRRSAAYLQHQRLHTGEKLYKCKECWKAFGCRS

LFIVHQRIHTGEKPYQCKECGKAFTQKIASIQHQRVHTGEKPYECKVCGKAFKWYGSF
VQHQKLHPVEKKPVKVLGPSLVSPQCSSPAIPPVLLQGSCSASAVAVPSLTFPHAVLI
PTSGNFFMLLPTSGIPSSSAQIVRVFQGLTPTVKPSPVILTPSSHSS"

ORIGIN

Query Match 20.6%; Score 413; DB 4; Length 8158;
Best Local Similarity 73.2%; Pred. No. 1.5e-55;
Matches 686; Conservative 0; Mismatches 195; Indels 56; Gaps 10;

```
Qy       1062 GTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACT 1121
             | |||||||||||| | ||||   || ||||| | ||||| |||| || |||  ||| | |
Db       8105 GGGAGTCACACAATTATTTTTATTTTCCAGTACCTATAACAGTTATGTTTAATCTATATT 8046

Qy       1122 GTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAA 1181
              |||||| |||||| |||||||| |||| ||| | ||   |||||| ||||||| ||||
Db       8045 ATAGTCTATTAAGTTTGCAATAGCATTATATCTGACAATGTACATACTTTAATTTAAAAA 7986

Qy       1182 TGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTT 1241
              |  |||||| |||||| |||||||| || || | ||  |||||||||        || ||
Db       7985 TATTTTATTGCTAAAAGATGCTAACGATAATCTAAGTCTTCAGTGAGT-----CTTTGTT 7931

Qy       1242 GCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT----------CGGGGGTCAGGTGCT 1290
              |||||  ||  |||||  || | || |||  | |              | | |    | ||||
Db       7930 GCTGGAAGACAGTCTTGCCTCACTGTTGATGGCTGCTGAATGATCGGTGTGGTGGTTGCT 7871

Qy       1291 GAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGCAATATC 1345
              |||| |||| |||||| | |   |||||| ||     ||||||  |||| ||||| |||
Db       7870 GAAGGTTAGTGTGGCTCTAGAGATTTCTTTAAATGAGACAACAAGGAAGGCTGCAACATC 7811

Qy       1346 AGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATT 1405
              |  |||||||||||||  ||| |||||| ||| |||  |||| |||| || |||||| ||
Db       7810 AACTGACTCTTCCTTTC-CGAAGGATTTATCTGTAGCATGTGTTGCTGTTGGATAGCCTT 7752

Qy       1406 TTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTT 1464
              |||  |||||| |||| |   || |||||||| | ||||||||||||| ||||||    ||
Db       7751 TTACCCACAGTGGAACATTCTTCAAAATTGAAGTCAATCCTCTCAAACTCTGCTGCTGTT 7692

Qy       1465 TAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACA 1524
              | |  | ||||||| || ||||||||| | ||| |||||||||||||||||||  |||||
Db       7691 TTATCAACTAAGTTTATGTAATATTCTAAGTCCTTTGTTGTCATTTCAACAATGGTCACA 7632

Qy       1525 GTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCAT 1584
              |  |||||||||||||||| ||||||||||       |  |    |  | |||| |  ||
Db       7631 GGATCTTCACCAGGAGTAGTTTCCATCTCAAGAAACCACTT----TTGTGGCTCTTATAT 7576

Qy       1585 AAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTC 1644
              ||||| || ||||||| |||||| |||||||||||||||||||||||||||||| ||||||  |
Db       7575 TAGAAGCAACTCCTCATTTGTTCACGTTTTATCATGAGATTGCAGCAATTCAGTCACATA 7516

Qy       1645 TTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTT 1704
              | ||||      | ||||| ||| || |||| ||||||| |||| ||||||||||  |||
Db       7515 TGCAGG-----TTCCACTTCTAAGTCTAGTTATCTTGCTATTTCCACCACATCTGCAGTT 7461

Qy       1705 CCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTC 1764
              |||| |||| |||||||||||||| ||| |  |||||||| |||||||   || |  ||
Db       7460 ACTTCGTCCACTGAAGTCTTGAACCCCTCAAGATCATCCATCAGGGTTGAGGTCAATGTC 7401

Qy       1765 TTCCAAATTCCTGTTAATATTTATATTTTGAC---CTCCCATGAATCATGAATGTTCTTA 1821
              |||  || |||||||||| || ||||||||||   ||||||||||||  ||||| ||||
Db       7400 TTCTGAACTCCTGTTAATGTTGATATTTTGACCTTCTCCCATGAATCACAAATGTCCTTA 7341

Qy       1822 ATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCA 1881
              |||| | || ||||| |||||||||||||| |||||||||||| |||||| || |||| | |
Db       7340 ATGGTATCTAGAATGATGAATCCTTTCCAGAAGGTTTTCAAGTTACTTTGTACAGAACAA 7281

Qy       1882 TCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAAT 1941
              |                  | ||| |||  | |||||||   |||||||| ||| |||
Db       7280 CC------------------ACTATGGCAGCCACAGCCTTACAAAATGTATTCCTTAAAT 7239

Qy       1942 AATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              ||||   |||||||||  |||||||||||||| ||||
Db       7238 AATA---CTTGAAAGTCAAAATTACTCCTTGACCCAT 7205
```

RESULT 3
HSM807294
LOCUS HSM807294 5937 bp RNA linear HTC 20-JAN-2005
DEFINITION Homo sapiens mRNA; cDNA DKFZp686B15223 (from clone DKFZp686B15223).
ACCESSION BX647150
VERSION BX647150.1 GI:34366178
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 5937)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.
CONSRTM The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2005) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Agowa (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp686B15223) is available at the RZPD Deutsches
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686B15223
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.
FEATURES Location/Qualifiers
source 1. .5937
/organism="Homo sapiens"
/mol_type="pre-RNA"
/db_xref="RZPD:DKFZp686B15223Q"
/db_xref="taxon:9606"
/clone="DKFZp686B15223"
/tissue_type="small intestine"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
/note="CD8 beta-chain glycoprotein, N-terminus truncated,
not fully spliced"

ORIGIN

Query Match 20.0%; Score 399.2; DB 4; Length 5937;
Best Local Similarity 70.7%; Pred. No. 2.4e-53;
Matches 728; Conservative 0; Mismatches 233; Indels 69; Gaps 12;

```
Qy        975 AGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAA 1034
              |||| ||||| ||||  ||||||     ||||  |||||| ||||||||||||  |||
Db        581 AGTATAGGCAAACCTCAGAGATATCTCAGGTTCAGTTCCAGACCACCACAATAAAGCAA- 639

Qy       1035 TATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGC 1094
                          ||||||||||||||  ||| |||    | |||||| |||||||||
Db        640 --------------ATATCACAATAAAGCCAGTTACATTCATTTTTTGGTTTCCCAGTGT 685

Qy       1095 ATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCT 1154
              | ||||||||| || |||  ||| |  |||||||  |||||||  ||| |      ||||||||
Db        686 ACATAAAAGTTATGTTTACCCTATAAGGTAGTCTACTAAGTGTACAAAGACATTATGTCT 745

Qy       1155 AAAAAAACACATAC--CTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212
              ||||||  |||  |||||||   |||||  |||||||| ||| |       | |||||
Db        746 GAAAAAAAAAGTACGTCTTAATCAAAAAATAATTTATTACCAAATACATGCTAAAATCAC 805

Qy       1213 TTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTT------------C 1260
              |||||  ||||||||||  |||| | |||||||  |||| ||  | ||
Db        806 CTGAGCTTTCAGTGAGTCATAATTTCTTTGCTGATGGAGGGCATCTTGCCTTGATGTCAA 865

Qy       1261 TTATTGATGACTGATC-GGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT 1319
              |   || |||||||| |   |   | |||  |||  | | |||||||| || |||| |
Db        866 TGGCTGCTGACTGATCAGAATGGTGGTTGCCAAAGGCTGGCGTGGCTGTGCCAATTTCGT 925
```

```
Qy       1320 AAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTC 1374
Db        926 AAAATAGGACAACAACAAATCTTGCCACATTGATTGACTCTTCCTTTCACAAAAGATTTA 985

Qy       1375 TCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATT 1434
Db        986 TCTGTAGCATGCGATGCTGTTTGATAGCATTTTACCTACAGTAGAACAGCATTCAAAATT 1045

Qy       1435 GGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGA 1493
Db       1046 GGAGTCACTCCTCTCAAATACTGACGCCTCTTTCTCAACTAAGTTTATGTAACATTCTCT 1105

Qy       1494 ATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTC 1553
Db       1106 ATCTTCTGTTGTCATTTCAACAATGTTCACACCATCTTCACCAGGAGTAGATTCCATTTC 1165

Qy       1554 ATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTT 1613
Db       1166 AAGAAACCACCT----TCTTTGCTCATCAGAGAGAATAAACTCCTCATGTGTT-AAGTTT 1220

Qy       1614 TATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAG 1673
Db       1221 GATCACGGGATTGCAACAATTCAGCCACGTCTTCAG-----ACTCCACTTCTATCTCTAG 1275

Qy       1674 TTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTC 1733
Db       1276 TTCTCTTGCTATTTCTACCATACTTGCAGTTACTTCCTCCACTGAAGTCTTGAACTCCTC 1335

Qy       1734 CAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTT 1793
Db       1336 CAACTCATCCAGGAGAATTGGAATCACCTGCTTCCAAACTCCTCCTAATGTTGGTATTTA 1395

Qy       1794 GAC---CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
Db       1396 TACCTGCTCCCATGAATCACAAATGTTCTTTGTTGCATCTAGAATGGTGAATCCTTTCCA 1455

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910
Db       1456 GAAGGTTTCACCCAGATCTA----------------TCAGAGGAATCATCATCTGTGGCA 1499

Qy       1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970
Db       1500 GCTCTAGCCTTACGAAATGTATTTCT----TAATAAGAATTGAAAGTCAAAATTACTCCT 1555

Qy       1971 TGATCCATTT 1980
Db       1556 TGGTCCATGT 1565
```

RESULT 4
BU616791

```
LOCUS       BU616791                707 bp    mRNA    linear   EST 23-SEP-2002
DEFINITION  UI-H-FH1-bfj-k-05-0-UI.s1 NCI_CGAP_FH1 Homo sapiens cDNA clone
            UI-H-FH1-bfj-k-05-0-UI 3', mRNA sequence.
ACCESSION   BU616791
VERSION     BU616791.1  GI:23283006
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Martin
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
            sequence: 20-707, >TIGGER1#DNA/MER2_type (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .707
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="UI-H-FH1-bfj-k-05-0-UI"
                      /tissue_type="Cell Line"
                      /dev_stage="Adult"
                      /lab_host="DH10B (Life Technologies)"
                      /clone_lib="NCI_CGAP_FH1"
                      /note="Organ: Chondrosarcoma; Vector: pT7T3-Pac
                      (Pharmacia) with a modified polylinker; Site_1: EcoR I;
                      Site_2: Not I; NCI_CGAP_FH1 is a normalized cDNA library
                      obtained from a cell line derived from grade I
                      chondrosarcoma tissue. The library was constructed and
                      normalized according to Bonaldo, Lennon and Soares, Genome
                      Research, 6:791-806, 1996. First strand cDNA synthesis was
                      primed with an oligo-dT primer containing a Not I site.
                      Double stranded cDNA was ligated to an EcoR I adaptor,
                      digested with Not I, and cloned directionally into
                      pT7T3-Pac vector. The oligonucleotide used to prime the
                      synthesis of first-strand cDNA contains a library tag
                      sequence that is located between the Not I site and the
                      (dT)18 tail. The sequence tag for this library is
                      AGAATCCGGC. The cell line was provided by Dr. James Martin
                      from the University of Iowa.
                      TAG_TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
                      Chondrosarcoma
                      TAG_LIB=UI-H-FH1
                      TAG_SEQ=AGAATCCGGC"
ORIGIN
```

Query Match 19.2%; Score 385; DB 5; Length 707;
Best Local Similarity 79.3%; Pred. No. 6.4e-51;
Matches 554; Conservative 0; Mismatches 115; Indels 30; Gaps 7;

```
Qy       1170 TTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC-ATTTGAGCATTCAGTGAG 1228
Db         18 TAAATTTAAAAATACTTTATTGCTAAAAAATGCTGATTATCAATCTGAGCCTTCGGTGAG 77

Qy       1229 TTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CG 1277
Db         78 TCGTACTCTTTTTGCTGGTGGAGGGTCTTGCCTCGAGGTTGATGGTTGCTGACTGGTCAG 137

Qy       1278 GGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGA 1332
Db        138 GGTGGTGGCTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCCTAAAATAAGACATCGGTGA 197

Qy       1333 AGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCT 1392
Db        198 AGCTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTGTAGCATGCGATGCT 257

Qy       1393 TTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAA 1451
Db        258 GTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAGTCAATCCTCTCAAA 317

Qy       1452 CCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTC 1511
Db        318 CCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGTTGTCATTTC 377

Qy       1512 AACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATC 1571
Db        378 AACAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGAAACCACTT----TC 433

Qy       1572 TTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCA 1631
Db        434 TTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATCATGAGATTGCAGCA 493
```

```
Qy       1632 ATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTAC 1691
             || |||| |  |||||||      ||  |||| |||||| |||||| ||||| |||| ||
Db        494 ATTCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCTTTTGCTATTTCCAC 548

Qy       1692 CACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGT 1751
             ||||||||   || |||||||| ||||||||||||| |||  ||||||| ||||||||
Db        549 CACATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAGTCATCAATGAGGGT 608

Qy       1752 TGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATC 1808
             |||||| |||||||||| |||||||||| || | | |||||   ||||||||||||
Db        609 GGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACCTCCTCCCATGAATC 668

Qy       1809 ATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
             | ||||||| ||||||||| ||  ||||||||||||||
Db        669 ACGAATGTTTTTAATGGCATCTAAAATGGTGAATCCTTT 707


RESULT 5
BU622317
LOCUS       BU622317                700 bp    mRNA    linear   EST 23-SEP-2002
DEFINITION  UI-H-FH1-bfu-h-08-0-UI.s1 NCI_CGAP_FH1 Homo sapiens cDNA clone
            UI-H-FH1-bfu-h-08-0-UI 3', mRNA sequence.
ACCESSION   BU622317
VERSION     BU622317.1  GI:23288532
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Martin
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
            sequence: 20-695, >TIGGER1#DNA/MER2_type (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FH1-bfu-h-08-0-UI"
                     /tissue_type="Cell Line"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_FH1"
                     /note="Organ: Chondrosarcoma; Vector: pT7T3-Pac
                     (Pharmacia) with a modified polylinker; Site_1: EcoR I;
                     Site_2: Not I; NCI_CGAP_FH1 is a normalized cDNA library
                     obtained from a cell line derived from grade I
                     chondrosarcoma tissue. The library was constructed and
                     normalized according to Bonaldo, Lennon and Soares, Genome
                     Research, 6:791-806, 1996. First strand cDNA synthesis was
                     primed with an oligo-dT primer containing a Not I site.
                     Double stranded cDNA was ligated to an EcoR I adaptor,
                     digested with Not I, and cloned directionally into
                     pT7T3-Pac vector. The oligonucleotide used to prime the
                     synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AGAATCCGGC. The cell line was provided by Dr. James Martin
                     from the University of Iowa.
                     TAG_TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
                     Chondrosarcoma
                     TAG_LIB=UI-H-FH1
                     TAG_SEQ=AGAATCCGGC"
ORIGIN

  Query Match             18.6%;  Score 373;  DB 5;  Length 700;
  Best Local Similarity   78.9%;  Pred. No. 5e-49;
  Matches  542;  Conservative    0;  Mismatches  115;  Indels   30;  Gaps    7;

Qy       1170 TTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC-ATTTGAGCATTCAGTGAG 1228
             | ||||| ||||| ||||||| |||||||||||| |  ||| || ||||| ||| |||||
Db         18 TAAATTTAAAAATACTTTATTGCTAAAAAATGCTGATTATCAATCTGAGCCTTCGGTGAG 77

Qy       1229 TTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CG 1277
             | ||| |||||||||||||||| ||||||  ||    | |||  | |            |
Db         78 TCGTACTCTTTTTGCTGGTGGAGGGTCTTGCCTCGAGGTTGATGGTTGCTGACTGGTCAG 137

Qy       1278 GGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGA 1332
             || |   | |||||||| || |||||||||||||| |||| ||||     ||| | ||||
Db        138 GGTGGTGGCTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCCTAAAATAAGACATCGGTGA 197

Qy       1333 AGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCT 1392
             || ||     || |||||||||||||||||||||||||||||||| |||  || ||||||
Db        198 AGCTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTGTAGCATGCGATGCT 257

Qy       1393 TTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAA 1451
             ||||||||||||||    |||||||||| |||||| |||||||| ||||||||||||||
Db        258 GTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAGTCAATCCTCTCAAA 317

Qy       1452 CCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTC 1511
             |||||||     || |  | ||||||| || |||||||| ||||| ||||||||||||
Db        318 CCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGTTGTCATTTC 377

Qy       1512 AACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATC 1571
             |||||| |||||||  |||||||||| ||||||||||||||        |  |    ||
Db        378 AACAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGAAACCACTT----TC 433

Qy       1572 TTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCA 1631
             ||||| |||||||||||   || ||||||||||| ||||||||||||||||||||||||||
Db        434 TTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATCATGAGATTGCAGCA 493

Qy       1632 ATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTAC 1691
             || |||| |  |||||||      ||  |||| |||||| |||||| ||||| |||| ||
Db        494 ATTCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCTTTTGCTATTTCCAC 548

Qy       1692 CACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGT 1751
             ||||||||   || |||||||| ||||||||||||| |||  ||||||| ||||||||
Db        549 CACATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAGTCATCAATGAGGGT 608

Qy       1752 TGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC---CATGAATC 1808
             |||||| |||||||||| |||||||||| || | | |||||||||||   ||||||||
Db        609 GGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACCTCCTCNCATGAATC 668

Qy       1809 ATGAATGTTCTTAATGGCACCTGGAAT 1835
             | ||||||| ||||||||| ||  |||
Db        669 ACGAATGTTTTTAATGGCATCTAAAAT 695


RESULT 6
CA430123
LOCUS       CA430123                690 bp    mRNA    linear   EST 07-NOV-2002
DEFINITION  UI-H-FH1-bfs-f-14-0-UI.s1 NCI_CGAP_FH1 Homo sapiens cDNA clone
            UI-H-FH1-bfs-f-14-0-UI 3', mRNA sequence.
ACCESSION   CA430123
VERSION     CA430123.1  GI:24792849
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 690)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Martin
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
            sequence: 20-690, >TIGGER1#DNA/MER2_type (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FH1-bfs-f-14-0-UI"
                     /tissue_type="Cell Line"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_FH1"
                     /note="Organ: Chondrosarcoma; Vector: pT7T3-Pac
                     (Pharmacia) with a modified polylinker; Site_1: EcoR I;
                     Site_2: Not I; NCI_CGAP_FH1 is a normalized cDNA library
                     obtained from a cell line derived from grade I
                     chondrosarcoma tissue. The library was constructed and
                     normalized according to Bonaldo, Lennon and Soares, Genome
                     Research, 6:791-806, 1996. First strand cDNA synthesis was
                     primed with an oligo-dT primer containing a Not I site.
                     Double stranded cDNA was ligated to an EcoR I adaptor,
                     digested with Not I, and cloned directionally into
                     pT7T3-Pac vector. The oligonucleotide used to prime the
                     synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AGAATCCGGC. The cell line was provided by Dr. James Martin
                     from the University of Iowa.
                     TAG_TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
                     Chondrosarcoma
                     TAG_LIB=UI-H-FH1
                     TAG_SEQ=AGAATCCGGC"
ORIGIN

  Query Match             18.0%;  Score 360.8;  DB 6;  Length 690;
  Best Local Similarity   79.0%;  Pred. No. 4.2e-47;
  Matches  539;  Conservative    0;  Mismatches  112;  Indels   31;  Gaps    8;

Qy       1170 TTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC-ATTTGAGCATTCAGTGAG 1228
              | ||||| ||||| ||||||| |||||||||||| |  ||| || ||||| ||| |||||
Db         18 TAAATTTAAAAATACTTTATTGCTAAAAAATGCTGATTATCAATCTGAGCCTTCGGTGAG 77

Qy       1229 TTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT----------CG 1277
              | ||| |||||||||||||||| ||||||  ||    | |||  | |           |
Db         78 TCGTACTCTTTTTGCTGGTGGAGGGTCTTGCCTCGAGGTTGATGGTTGCTGACTGGTCAG 137

Qy       1278 GGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGA 1332
              || |   | |||||||| || |||||||||||||| |||| ||||     ||| | ||||
Db        138 GGTGGTGGCTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCCTAAAATAAGACATCGGTGA 197

Qy       1333 AGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCT 1392
              || ||     || ||||||||||||||||||||||||||||||| |||  || ||||||
Db        198 AGCTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTGTAGCATGCGATGCT 257

Qy       1393 TTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAA 1451
              ||||||||||||||    ||||||||| ||||||| |||||||| ||||||||||||||
Db        258 GTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAGTCAATCCTCTCAAA 317

Qy       1452 CCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTC 1511
              |||||||     || |  | ||||||| || ||||||||| ||||| ||||||||||||
Db         318 CCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGTTGTCATTTC 377

Qy       1512 AACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATC 1571
              |||||| |||||||  |||||||||||| ||||||||||||||||        | |    ||
Db         378 AACAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGAAACCACTT----TC 433

Qy       1572 TTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCA 1631
              ||||| |||||||||||  || ||||||||||| ||||||||||||||||||||||||||
Db         434 TTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATCATGAGATTGCAGCA 493

Qy       1632 ATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTAC 1691
              || |||| |  ||||||||     ||  |||| |||||| |||||| ||||| |||| ||
Db         494 ATTCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCTTTTGCTATTTCCAC 548

Qy       1692 CACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGT 1751
              ||||||||   || |||||||| |||||||||||||| |||  ||||||| ||||||||
Db         549 CACATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAGTCATCAATGAGGGT 608

Qy       1752 T-GGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAAT 1807
              | |||||| |||||||||||| |||||||||| || | | |||||    |||||||||||
Db         609 TGGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACCTCCTCCCATGAAT 668

Qy       1808 CATGAATGTTCTTAATGGCACC 1829
              || ||||||| ||||||||| |
Db         669 CACGAATGTTTTTAATGGCATC 690


RESULT 7
HSM806813
LOCUS       HSM806813               9266 bp    mRNA    linear   HTC 22-SEP-2004
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686M21107 (from clone DKFZp686M21107).
ACCESSION   BX640738
VERSION     BX640738.1  GI:34364830
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 9266)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp686M21107) is available at
            the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in
            Berlin, Germany. Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686M21107
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .9266
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp686M21107Q"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686M21107"
                     /tissue_type="fetal kidney"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="fetal"
                     /note="hypothetical protein, C-terminus elongated"
     gene            1. .9266
                     /gene="DKFZp686M21107"
```

```
     CDS             406. .3903
                     /gene="DKFZp686M21107"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAE45851.1"
                     /db_xref="GI:34364831"
                     /db_xref="UniProt/TrEMBL:Q6N021"
                     /translation="MEQDRTNHVEGNRLSPFLIPSPPICQTEPLATKLQNGSPLPERA
                     HPEVNGDTKWHSFKSYYGIPCMKGSQNSRVSPDFTQESRGYSKCLQNGGIKRTVSEPS
                     LSGLLQIKKLKQDQKANGERRNFGVSQERNPGESSQPNVSDLSDKKESVSSVAQENAV
                     KDFTSFSTHNCSGPENPELQILNEQEGKSANYHDKNIVLLKNKAVLMPNGATVSASSM
                     EHTHGELLEKTLSQYYPDCVSIAVQKTTSHINAINSQATNELSCEITHPSHTSGQINS
                     AQTSNSELPPKPAAVVSEACDADDADNASKLAAMLNTCSFQKPEQLQQQKSVFEICPS
                     PAENNIQGTTKLASGEEFCSGSSSNLQAPGGSSERYLKQNEMNGAYFKQSSVFTKDSF
                     SATTTPPPPSQLLLSPPPPLPQVPQLPSEGKSTLNGGVLEEHHHYPNQSNTTLLREVK
                     IEGKPEAPPSQSPNPSTHVCSPSPMLSERPQNNCVNRNDIQTAGTMTVPLCSEKTRPM
                     SEHLKHNPPIFGSSGELQDNCQQLMRNKEQEILKGRDKEQTRDLVPPTQHYLKPGWIE
                     LKAPRFHQAESHLKRNEASLPSILQYQPNLSNQMTSKQYTGNSNMPGGLPRQAYTQKT
                     TQLEHKSQMYQVEMNQGQSQGTVDQHLQFQKPSHQVHFSKTDHLPKAHVQSLCGTRFH
                     FQQRADSQTEKLMSPVLKQQLNQQASETEPFSNSHLLQHKPHKQAAQTQPSQSSHLPQ
                     NQQQQQKLQIKNKEEILQTFPHPQSNNDQQREGSFFGQTKVEECFHGENQYSKSSEFE
                     THNVQMGLEEVQNINRRNSPYSQTMKSSACKIQVSCSNNTHLVSENKEQTTHPELFAG
                     NKTQNLHHMQYFPNNVIPKQDLLHRCFQEQEQKSQQASVLQGYKNRNQDMSGQQAAQL
                     AQQRYLIHNHANVFPVPDQGGSHTQTPPQKDTQKHAALRWHLLQKQEQQQTQQPQTES
                     CHSQMHRPIKVEPGCKPHACMHTAPPENKTWKKVTKQENPPASCDNVQQKSIIETMEQ
                     HLKQFHAKSLFDHKALTLKSQKQVKVEMSGPVTVLTRQTTAAELDSHTPALEQQTTSS
                     EKTPTKRTAASVLNNFIESPSKLLDTPIKNLLDTPVKTQYDFPSCRCVGKCQKCTETH
                     GVYPELANLSSDMGFSFFF"
ORIGIN

  Query Match           18.0%;  Score 360.6;  DB 4;  Length 9266;
  Best Local Similarity   67.6%;  Pred. No. 2.6e-47;
  Matches  783;  Conservative    0;  Mismatches  284;  Indels   92;  Gaps   16;

Qy        868 CTCCTAATTTTTATGCTATGTCTATAATACATTAGGTTTGTTGTTGTTTGTTTTTACCTT 927
              ||||||| |||    |    | | ||||||| ||  |||||| ||| | || |  ||
Db       4333 CTCCTAAAGTTTCCTTTTCTCCATTTATACATTTGGAATGTTGTGATTTATATTCATTTT 4392

Qy        928 ATTGCTGTTGGCTGGGGTCAGCAAACATTTTCTGTAAAGGGCTAGATAGTACAGGCATAC 987
                | |  ||  ||    | |     ||| || |  |     |  | ||  ||||||||||
Db       4393 GATTCCCTTTTCT---CTAAAATTTCATCTTTTTGATTAAGAAATATGATACAGGCATAC 4449

Qy        988 CTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAAGAAGTG 1047
              ||  |||||| |||||||||| |||||||||| | || |                  ||
Db       4450 CTCAGAGATATTGTGGGTTTGGCTCCATACCACAATAAAA----------------TG 4491

Qy       1048 GATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTT 1107
                |||| |||||||||  |||    |      ||||||  ||| || || || |||||||| |
Db       4492 AATATTACAATAAAGCAAGTTGTAAGGACTTTTTGGTTTCTCACTGTATGTAAAAGTTAT 4551

Qy       1108 GCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATA 1167
                |||||||  ||||||    |  ||||||||||||||  || ||||||||||     |||
Db       4552 TTATATACTATACTGTAACATACTAAGTGTGCAATAGCATTGTGTCTAAAAA-----ATA 4606

Qy       1168 CCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGA 1227
                |   ||  |||    ||  |    |||||||| |||||  || || || || |||||
Db       4607 TATACTTTAAAAATAATTTATTGTTAAAAAAATGCCAACAATTATCTGGGCCTTTAGTGA 4666

Qy       1228 GTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGGTC---- 1283
              ||  |||||||||||||||||| |||| |   | | |  |||  | |  ||  |
Db       4667 GTGCTAATCTTTTTGCTGGTGGAGGGTCGTGCTTCAGTATTGATCGCTGTGGACTGATCA 4726

Qy       1284 ----------------AGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAAC 1327
                             ||||      | |    |||  ||||||| ||||||||| ||
Db       4727 TGGTGGTAGTTGCTGAAGGTTGCTGGGATGGCTGTGTGTGTGGCAATTTCTTAAAATAAG 4786

Qy       1328 AGTGAAG---ATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGT 1384
                |    ||   | |||  ||||| |||| | |||| ||||  |||||||||| |||  |
Db       4787 ACAACAGTGAAGTGCTGTATCAATTGATTTTTCCATTCACAAAAGATTTCTCTGTAGCAT 4846

Qy       1385 GTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATC 1443
              |  ||||| |||||||||||||| |  ||||| |||| |||||||||||||||| ||| ||
Db       4847 GCAATGCTGTTTGATAGCATTTAACCCACAGCAGAATTTCTTTGAAAATTGGACTCAGTC 4906

Qy       1444 CTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTT 1503
              |||||||| |||     || |  | |||||||  | ||||||||||||||| |||||
Db       4907 CTCTCAAACTGTGCTGCTGCTTTATCAACTAAGTTTTTGTAATATTCTGAATCCTTTGTT 4966

Qy       1504 GTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAG 1563
              |||||||| || |||  |||| ||||||   | |||| ||||||||||
Db       4967 GTCATTTCAGCAGTTT--ACAGCATCTTCATTGGAAGTATATTCCATCTC---------- 5014

Qy       1564 ATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGA 1623
              |   | |||||| |||||||||||||| ||      | |||  |||||||| |||||| |
Db       5015 AAACATTCTTTGTTCATCCATAAGAAGCAA-----CTTCTTATCAAGTTTTTTCATGACA 5069

Qy       1624 TTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCT 1683
              |||||| ||  ||| |   ||||||||      ||  |||| ||||||  ||||||||||
Db       5070 TTGCAGTAACTCAGCCCCATCTTCAGG-----CTCTACTTCTAATTCTGGTTCTCTTGCT 5124

Qy       1684 GTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCC 1743
                 ||| || ||||||  ||  | | |||||  ||||||||||||  ||| |||| ||||
Db       5125 ACATCTCCCTCATCTGCAGTGACCT-CTCCACGGAAGTCTTGAACTCCTCAAAGTAATCC 5183

Qy       1744 ATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCC 1800
              |||||||||||||||| ||||||   ||| ||||||||||  || |||| ||||   |||||
Db       5184 ATGAGGGTTGGAATCAACTTCT---AAACTCCTGTTAATGTTGATATATTGACCCCCTCC 5240

Qy       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860
              ||||||| |||||||||||||||  |  ||  ||||||||  ||||||||| |||| ||||
Db       5241 CATGAATTATGAATGTTCTTAATAACTTCT-AAATGGTGATACCTTTCCAGAAGGCTTTC 5299

Qy       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920
              ||| ||||| | || ||||||||          |       |    || ||| | ||| ||
Db       5300 AATGTACTTTGCCCGGATCCATC----------AGAAGACTATCTTGGCAGCTGTAGACT 5349

Qy       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980
                |   ||| ||| |||| ||| ||||| |||||||||  ||| ||||||||  ||||||
Db       5350 AA--CAATATATCTCTTAAATGATAAGACTTGAAAGTCAAAAGTACTCCTTAATCCATAG 5407

Qy       1981 TCTGCAAAATAGATGTTGT 1999
                 ||||| |||  |||||||
Db       5408 GCTGCAGAATCAATGTTGT 5426


RESULT 8
CB957195/c
LOCUS       CB957195                778 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  AGENCOURT_13666972 NIH_MGC_184 Homo sapiens cDNA clone
            IMAGE:30353361 5', mRNA sequence.
ACCESSION   CB957195
VERSION     CB957195.1  GI:30213312
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 778)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM153  row: p  column: 10
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1. .778
                     /organism="Homo sapiens"
```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30353361"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_184"
/note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); Library is oligo-dT primed and directionally cloned. cDNA was prepared from a glandular pool of tissues from thyoid, parathyroid, adrenal, cortex and pineal gland. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.38 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 18.0%; Score 360.4; DB 6; Length 778;
Best Local Similarity 76.5%; Pred. No. 4.8e-47;
Matches 525; Conservative 0; Mismatches 141; Indels 20; Gaps 6;

```
Qy      1320 AAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       676 ACAACAACAATGAAGTTTGCTGCATCGACTGATTCTCCCTTTCACAAAAAATTTTTCTGT 617

Qy      1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-T 1438
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       616 AGCATGCAGTCCTATTTGGTAGCATTTTACCCACAGCAGAACTTCTTTCAAAATTAGAGT 557

Qy      1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAA--TATTCTGAATC 1496
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       556 CGATACTGTCAAACTCTGCCACTGCTTTATCAACTAAGTTTATATAATGTACTCTAAGTC 497

Qy      1497 CATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATT 1556
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       496 CTTTGTTGTCATTTCAACCATGTTAACAGCATCTTCACCAGGCGTAGATTCCATCTCAAG 437

Qy      1557 TCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTAT 1616
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       436 AAACCACTT----TCTTTGTTCATCCCTAAGAAGCAACTCCTCATCCTTTCAAATTTTAT 381

Qy      1617 CATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTC 1676
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       380 CATGAGATTGCAGGAATTCAGTCACATCTTCAGG-----CTCCACTTCTAATTTGAGTTA 326

Qy      1677 TCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAA 1736
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       325 TCTTACTATTTCCACCACATCTGCAGTTATTTCCTCCACTGAAGTCTTGAACCCCTCAAA 266

Qy      1737 GTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTG-- 1794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       265 GTTATCCATGAAGGTTGTAATCAACTTCTTCCAAATTCCTGTTAATGTTGATATTTTGAC 206

Qy      1795 -ACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAA 1853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       205 CACCTCCCATGAATCACAAATATTCCTAATGGTTATCTAGAATTAGAATCCTTTCCAGAA 146

Qy      1854 GGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTT 1913
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       145 AGTTTTCAATTGACTTTGTCCAAATGAAT-----CAGAAGAATCACTATCTATGGCAGCT 91

Qy      1914 ATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGA 1973
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 ATAGCCTTACAAAATGTATTTCTTCAATAGTAAGACTTGAAAGTGAAAATTGCTCCTTGA 31

Qy      1974 TCCATTTTCTGCAAAATAGATGTTGT 1999
             |||||||||||||||||||||||||
Db         30 TCCATGGGCTGAGGAAAAGATGCTGT 5
```

RESULT 9
BG113167/c
LOCUS BG113167 994 bp mRNA linear EST 30-JAN-2001
DEFINITION 602283837F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4371176 5', mRNA sequence.
ACCESSION BG113167
VERSION BG113167.1 GI:12606673
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 994)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10029 row: j column: 09
High quality sequence stop: 637.

FEATURES Location/Qualifiers
source 1. .994
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4371176"
/tissue_type="osteosarcoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_86"
/note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.533 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 17.6%; Score 351.6; DB 2; Length 994;
Best Local Similarity 77.5%; Pred. No. 1.1e-45;
Matches 535; Conservative 0; Mismatches 129; Indels 26; Gaps 8;

```
Qy      1257 TTTCTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       675 TTGATGGTTCTTGACTGATCAGGGTGTGGGCCTCTGAAAGTTGGGGTGCTGTTGCAATTT 616

Qy      1317 CTTAAA----ACAACAGTGAAGATTGCAATATCAGTT---GACTCTTCCTTTCATGAAAG 1369
             ||||||    |||||||||||||||||||||||||||   |||||||||||||||||||
Db       615 CCTAAATAAGACATCGGTGAAGCTTTTCGCATGAGGTTGACTCTGTCCTGTTCATGAAAG 556

Qy      1370 ATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGA 1429
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       555 ATTTCTCTGTAGCATGCGATGCTGTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCA 496

Qy      1430 AAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATAT 1488
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       495 AAATTGGAGTCAATCCTCTCAAACCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATAT 436

Qy      1489 TCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCC 1548
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       435 TCTAAATCCTTTGTTGTCATTTCAACAATGTTCACAGCATCTTCACCAGGTGTAGATTCC 376

Qy      1549 ATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCA 1608
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       375 ATCTCAAGAAACCACTT----TCTTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAA 320

Qy      1609 AGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAAT 1668
```

```
Db        319 AGTTTTATCATGAGA-TGCAGCAATTCAGTTACATCTTCAGGCTCTACTTC-----TAAT 266

Qy       1669 TCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAAC 1728

Db        265 TCTAGTTCTTTTGCTATTTCCACCACATCTGCACTTACTTCCTCCACTGAAGTCTTGAAC 206

Qy       1729 CTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTAT 1788

Db        205 CCCTCATAGTCATCAATGAGGGTGGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGAC 146

Qy       1789 ATTTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCT 1845

Db        145 AATTTGACCTCCTCCCATGAATCACGAATGTTTTTAATGGCATCTAAAATGGTGAATCCT 86

Qy       1846 TTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAA 1905

Db          85 TTCCAGAAGGTTTTCAATTTGCTTTGCCCAGATCCAT-----CAGAGACATCACTATCCA 31

Qy       1906 TGCCAGTTATAGCCTTATGGAATGTATTTC 1935

Db          30 TGGCAGCTAGAGCCTTATGAAATGTATTTC 1
```

```
RESULT 10
AU120614
LOCUS       AU120614                814 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU120614 HEMBB1 Homo sapiens cDNA clone HEMBB1001060 5', mRNA
            sequence.
ACCESSION   AU120614
VERSION     AU120614.1  GI:10935849
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 814)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .814
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HEMBB1001060"
                     /tissue_type="whole embryo, mainly body"
                     /dev_stage="embryo, 10 weeks"
                     /clone_lib="HEMBB1"
                     /note="Vector: pME18SFL3"
ORIGIN

  Query Match             17.1%;  Score 341.2;  DB 1;  Length 814;
  Best Local Similarity   73.8%;  Pred. No. 5e-44;
  Matches  563;  Conservative    0;  Mismatches  169;  Indels   31;  Gaps    9;

Qy       1055 CAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATA 1114

Db          3 CAATAAAACAAGTCATATGAATGTTTTGGTTTCCCAGTGCATAT-AAAGTTTTGTTTACA 61

Qy       1115 CTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAA---CACATACCTT 1171

Db          62 CTATACTGTAGTCTATT-TGGGTGCAATGGCATTATGTCTAAAAACAATGTAAATACCTT 120

Qy       1172 AATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT--TTGAGCATTCAGTGAGT 1229

Db         121 AATTAAAAAATACTATATTGCTAAAAGATGCTACCAATCATCGCTGAGCCTTCAGTGAGT 180

Qy       1230 TGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTT----------ATTGATGACTGATCGGG 1279

Db         181 AGTAATCTTTATGCTGGCAGAGGGCCTTGCCTTGATGTTGATAGCTGCTGACTGATCAGA 240

Qy       1280 GGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGC 1339

Db         241 ATGTGTTTGGTGAAGCTTGGGCTGGCTGCAGCAATTTTGTAAGACAACAATGAAGTTTGC 300

Qy       1340 AATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGAT 1399

Db         301 CTCCTTGATGGATTCTTCCTTTCACGAAAGATTTCTCCATAGCATGCCATGCTGTTTGAT 360

Qy       1400 AGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCT 1458

Db         361 AGGGTTTTACCCACAGTAAATCTTCTTTCAAAATTGGAGTTGATCCTCTCGAATCCTGCT 420

Qy       1459 CTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAAC-AAT 1517

Db         421 GCTGCCTTATCAACAAAGTTTATAGAATTTTCTAAACCCTTTGTTGTCATTTCAACAAAT 480

Qy       1518 TTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCT 1577

Db         481 TTTCACAACATCTTCGCCAGGAGTAGATTCCAGGTCAAGAAACCACTT----TCTTTATT 536

Qy       1578 CATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAG 1637

Db         537 TAACCATAAGAAGCAACTGCTCATTCATTGAAGTTTTATCATGAGACTGCAGCAAGTCAG 596

Qy       1638 TCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATC 1697

Db         597 GCACATCTTGAGGCTCCACTT--------ATTCTAATTCTCTTGCTGGTGCTACCATATC 648

Qy       1698 TGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAAT 1757

Db         649 TGCAGTTACTTCTTCCTCTGAAGTCTTGAACCCCTCAAAGTTATTCATGAGGGTTAAAAT 708

Qy       1758 CGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800

Db         709 CAACTTCTTNCAAACTTCTGGTAATGGTGCTATTTTTACTTCC 751
```

```
RESULT 11
BX955959/c
LOCUS       BX955959                719 bp    mRNA    linear   EST 01-MAR-2004
DEFINITION  DKFZp781A1775_r1 781 (synonym: hlcc4) Homo sapiens cDNA clone
            DKFZp781A1775 5', mRNA sequence.
ACCESSION   BX955959
VERSION     BX955959.1  GI:43439531
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
            Wellenreuther,R., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
```

This is the 5' sequence of the clone insert .
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de; sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
No s1 sequence available.
This clone (DKFZp781A1775) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES Location/Qualifiers
source 1. .719
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp781A1775"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="781 (synonym: hlcc4)"
/note="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"

ORIGIN

Query Match 16.8%; Score 335.6; DB 5; Length 719;
Best Local Similarity 75.2%; Pred. No. 3.9e-43;
Matches 533; Conservative 0; Mismatches 145; Indels 31; Gaps 8;

```
Qy       1195 AAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGT 1254
             |||||||||| |||||| | ||  |||||  || ||||||  || |||| |||||||||
Db        719 AAAAATGCTAATAATCATCTAAGTCTTCAGCAAGGTGTAAT-ATTCTGCTAGTGGAAGGT 661

Qy       1255 CTTTTCTTATTGATGACTGATC-----------GGGGGTCAGGTGCTGAAGCTTAGGGTG 1303
             ||    | | || |||  | |            ||| |   | |||||||| || | |||
Db        660 CTCACTTCAATGNTGATGGCTGCAAACTGATCAGGGCGGTGGTTGCTGAAGGTTGGAGTG 601

Qy       1304 GCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCC 1358
             ||||||||| ||    |||      |||||| | ||| |||| | ||   | ||||||| |
Db        600 GCTGTGGCAATTGAAAAAAATAAGACAACAATTAAGTTTGCCACATTGATGGACTCTTTC 541

Qy       1359 TTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAG 1418
             |||  |||||||| ||||| |||   | || ||||||||||||  |||||  || || ||
Db        540 TTTTGTGAAAGATATCTCTGTAGCATTTGGTGCTTTTTGATAGTGTTTTACCCATAGCAG 481

Qy       1419 AACTTCTTTGAAAATT-GGATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGT 1477
             |||||||||| ||||| || |||||||||||||||||   | |||||||| ||| ||||
Db        480 AACTTCTTTGGAAATTGGGGTCAATCCTCTCAAACCTGCCGCTGCTTTATCAA-CTAACC 422

Qy       1478 TAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAG 1537
             | ||  |||||||| ||||  || |||| |||||||||  |||||||  |||||||||
Db        421 TTATGAAATATTCTAAATCTTTTATTGTGATTTCAACAACGTTCACAGCATCTTCACCAG 362

Qy       1538 GAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCC 1597
             ||| ||||||||||    ||    ||       | |||||||||| |||||| || |||
Db        361 GAGCAGATTCCATC----TTAAGAAGCCACTTTATTTGCTCATCCAGAAGAAGCAACTCC 306

Qy       1598 TCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACT 1657
             ||||| ||||| ||||| |||||||||||||||||| ||||||  |||||| |      |
Db        305 TCATCCGTTCATGTTTTCTCATGAGATTGCAGCAATTCAGTCACATCTTCAAG-----AT 251

Qy       1658 TCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTG 1717
              ||||| |||| | |||||||||||| ||||||| |||||||   |  |||||||| |
Db        250 CCACTTCTAATCCTAGTTCTCTTGCTATTTCTACGACATCTGCATTGACTTCCTCCACTA 191

Qy       1718 AAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTG 1777
             ||  |||||| | ||| || |||| |||||||||| ||    ||| |||||||| || ||
Db        190 AAACCTTGAATCCCTCAAAATCATTCATGAGGGTTGAAAGTAACTCCTTCCAAACTCTTG 131

Qy       1778 TTAATATTTATATTTTGACCTC---CCATGAATCATGAATGTTCTTAATGGCACCTGGAA 1834
             ||||| || ||||||| |||||    ||||||||| ||| ||||||||||||||    |||
Db        130 TTAATGTTGATATTTTCACCTCCTTCCATGAATCGTGACTGTTCTTAATGGCATACAGAA 71

Qy       1835 TGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATC 1883
             |||| ||| ||||||| |||||||||||||||||| | ||| |||||||
```

```
Db         70 TGGTAAATTCTTTCCAGAAGGTTTTCAATTTACTTTGCCCATATCCATC 22
```

RESULT 12
CF242986/c
LOCUS CF242986 907 bp mRNA linear EST 05-AUG-2003
DEFINITION AGENCOURT_15115767 NIH_MGC_192 Homo sapiens cDNA clone IMAGE:30512711 5', mRNA sequence.
ACCESSION CF242986
VERSION CF242986.1 GI:33446198
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 907)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Agencourt
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM573 row: g column: 24
High quality sequence stop: 671.

FEATURES Location/Qualifiers
source 1. .907
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30512711"
/tissue_type="Brain glioblastoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_192"
/note="Vector: pExpress1; Site_1: SmaI; Site_2: NotI; The library was constructed by reverse transcription of 1 ug mRNA using the oligo dT primer GCGGCCGCCC(T)20 and an RNaseH + MMLV reverse transcriptase. Second strand synthesis was carried out by standard methods. The cDNA was size selected by agarose gel for > 1.2 kb, digested with Not I and directionally cloned into the vector Express-1 at the SmaI/NotI sites. DNA from the primary library was used for in vitro transcription from the T7 promoter to produce biotinylated RNA transcripts. These biotinylated transcripts, along with blocking oligos to the poly-A, multiple cloning site and primer regions, were hybridized with single stranded circles produced by phagmeid production from the primary library to a Cot value of 10-20. Strepavidin/phenol extraction was utilized to remove DNA:RNA hybrids leaving un-hybridized single stranded circles which were repaired by primer extension and transformed back into E. coli resulting in the normalized library. Average insert size 2.0 kb. 3' linker/adaptor sequence GCGGCCGCCC(T)20. This libary was constructed by Agencourt Bioscience."

ORIGIN

Query Match 16.8%; Score 335.2; DB 6; Length 907;
Best Local Similarity 73.5%; Pred. No. 4.3e-43;
Matches 545; Conservative 0; Mismatches 164; Indels 32; Gaps 8;

```
Qy       1156 AAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTG 1215
             || |  |  ||  |||||||| |||  | |||||| |||||||||| |||  ||||| ||
Db        728 AACACGCCAATTTCTTAATTTAAAATAGTTTTATTGCTAAAAAATGTTAATGATCATCTG 669
```

```
Qy       1216 AGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTT----------TTCTTAT 1264
             ||| ||||| |||| ||||||||||||||| |||| ||||||           | |
Db        668 AGCCTTCAGAGAGTCTTAATCTTTTTGCTGATGGAGGGTCTTGCCTCCATGGCTGATGGC 609

Qy       1265 TGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-- 1322
             || ||||||||| ||    |||| ||    | | |    ||| || | ||||
Db        608 TGCTGACTGATCAGGATGGAGGTTGCTAAATGATGAGGCGGCTGTGCAATTGCATAAAAT 549

Qy       1323 ---ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
                |||||||| ||| ||| |||| |||||||||| |||||  ||| | ||||||
Db        548 AAGACAACAGTAAAGTCTGCCTCATCAACTGACTCTTCTTTTCACCAAAAACTTCTCTGT 489

Qy       1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTG-GAT 1438
             ||  ||   || | | |||||| ||||||| ||||||||||||||||| |||| || | |
Db        488 AGCATGCAGTGTTGTGTGATAGTATTTTATCCACAGTAGAACTTCTTTCAAAACTGTGGT 429

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
             | ||||||||| ||||    || |||| ||| ||||||| |  | ||||||| |||||
Db        428 CTATCCTCTCAGACCCGCTGCTCCTTTTTCAA-CTAAGTTTAAGTCATATTCTAAATCCC 370

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
             | |||| || || || |||  |||||| ||||||||||||| |||||||||  ||||
Db        369 TCGTTGCCAGTTGAATAATGCTCACAGCATCTTCACCAGGACTAGATTCCACCTCAAGAA 310

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
                | |    |||||||||| ||| | ||| || ||||||||  ||||||||||   |
Db        309 ACCACGT----TCTTTGCTCATTCATTAAAAGCAACTCCTCATCCATTCAAGTTTT---A 257

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
             |||||||||||||||| |||| |  |||||||     || ||||| |||||| |||||||
Db        256 TGAGATTGCAGCAATCCAGTTACATCTTCAGG-----CTGCACTTCTAATTCTAGTTCTC 202

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
             ||||| |||| |||| || ||| ||| |||||| || ||||| |||||||| ||  ||||
Db        201 TTGCTATTTCCACCATATGTGTAGTTACTTCCTACACTGAAGGCTTGAACCCCTTAAAGT 142

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
             |||||||||||||||||||| |||||||||||| |||||||||| || | | ||||||||
Db        141 CATCCATGAGGGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACACTTTGACCT 82

Qy       1799 CC--CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGT 1856
                |  | |||||||||| ||||||| ||||| || ||||||   |||||||||  ||| |
Db         81 TCTTCCTGAATCATGAGTGTTCTTCCTGGCATCTAGAATGGCACATCCTTTCCCTAAGTT 22

Qy       1857 TTTCAATTTACTTAGTCCAGA 1877
             |||| |||||||| | ||| |
Db         21 TTTCTATTTACTTTGCCCANA 1


RESULT 13
AQ320818/c
LOCUS       AQ320818                603 bp    DNA     linear   GSS 06-MAY-1999
DEFINITION  RPCI11-92B20.TV RPCI-11 Homo sapiens genomic clone RPCI-11-92B20,
            genomic survey sequence.
ACCESSION   AQ320818
VERSION     AQ320818.1  GI:4053552
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-92B20.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7534987"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-92B20"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             16.5%;  Score 330.6;  DB 9;  Length 603;
  Best Local Similarity   77.1%;  Pred. No. 2.5e-42;
  Matches  471;  Conservative    0;  Mismatches  124;  Indels   16;  Gaps    5;

Qy       1343 ATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGC 1402
             ||  || ||||||| |||||||||| || || ||| |||| || |||||| ||||||| |
Db        601 ATTGGTGGACTCTTACTTTCATGAATGACTTTTCTGTAGTATGAGATGCTGTTTGATACC 542

Qy       1403 ATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACCCTGCTCTGC 1462
             || |||  |||||||||||||||||| |||| ||| ||||| |||||||||  ||
Db        541 ATCTTACCCACAGTAGAACTTCTTTCAAAAGTGG-TCAATTCTCTCAAACCTGGCATCTG 483

Qy       1463 TTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCA 1522
              || |  | ||||||| || ||||||| | ||||| |||||| ||||||| ||| ||||
Db        482 CTTTATCAGCTAAGTTTATGTAATATTGTTAATCCTTTGTTGACATTTCAATAATGTTCA 423

Qy       1523 CAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCC 1582
             |||  |||||| ||||||||| | ||||| |        || |    ||||||||||||
Db        422 CAGCATCTTCACTAGGAGTAGACTTCATCTTAAGAAACCAGTT----TCTTTGCTCATCC 367

Qy       1583 ATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATG 1642
             |||||||| || | |||||||||||| ||||||||||||||||||||||||| ||||||
Db        366 ATAAGAAGTAACTTCTCATCTGTTCATGTTTTATCATGAGATTGCAGCAATTCAGTCACA 307

Qy       1643 TCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGG 1702
             |||||||||  ||||||     |||||| |||||||| ||| ||| ||| ||||||| |
Db        306 TCTTCAGGCTCCACTTC-----TAATTCTAGTTCTCTCGCTATTTTTACTACATCTGCAG 252

Qy       1703 TTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACT 1762
             || |||| |||| |||||||||||||| ||  | ||||  ||||||   ||||||| |||
Db        251 TTACTTCTTCCACTGAAGTCTTGAACCCCTTAACGTCAATCATGAGATCTGGAATCAACT 192

Qy       1763 TCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATCATGAATGTTCT 1819
             ||||||| | |||||||||| | ||||||||   |||||||| || |||| |||||||
Db        191 TCTTCCAGACTCCTGTTAATGTGGATATTTTGACCTCCTCCCATAAACCATGCATGTTCT 132

Qy       1820 TAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATC 1879
             || |||||  | |||||||  ||| ||||  | |||| |||||||| |||| |  |||||
Db        131 TAGTGGCATGTAGAATGGCAAATTCTTTTTAGAAGGCTTTCAATTAACTTTGCACAGATG 72

Qy       1880 CATCCATC---CAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCT 1936
             |    |||   |||||||  ||||  |  || ||| |||| |||||||   |||||||||
Db         71 CCCAGATCTATCAGAGGAATCACTACCTGTGGCAGCTATAACCTTATGATTTGTATTTCT 12

Qy       1937 TCAATAATAAG 1947
             | ||| ||| |
Db         11 TAAATTATAGG 1
```

```
RESULT 14
BZ603609
LOCUS       BZ603609                687 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHAAV41TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-6G10, genomic survey sequence.
ACCESSION   BZ603609
VERSION     BZ603609.1  GI:31512071
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-6G10"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN

  Query Match             16.5%;  Score 330;  DB 9;  Length 687;
  Best Local Similarity   76.5%;  Pred. No. 3e-42;
  Matches  519;  Conservative    0;  Mismatches  125;  Indels   34;  Gaps    8;

Qy       1333 AGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCT 1392
             || |||| | |||| ||||||||| |||||||||| |||||||| |||  | |||||||
Db             1 AGCTTGCCACATCAATTGACTCTTTCTTTCATGAAAAATTTCTCTGTAGCATATGATGCT 60

Qy       1393 TTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAA 1451
              ||||||||||||||  || |||||||  ||||  | |||||||  |||||||||||||
Db            61 GTTTGATAGCATTTTACCCAAAGTAGAAAATCTTGTACAATTGGAGTCAATCCTCTCAAA 120

Qy       1452 CCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTC 1511
              |||||||     ||||      |||||  ||  ||||||||  |||||  ||||||||||
Db           121 CCCTGCTGCTGCTTAA------AAGTTTATGGAATATTCTAAATCCTATGTTGTCATTTT 174

Qy       1512 AACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATC 1571
              | | || |||||||  | |||| ||||||||||||||| ||| |      ||       ||
Db           175 AGCCATGTTCACAGCATTTTCATCAGGAGTAGATTCCGTCTTA-----AGACGCATTTTC 229

Qy       1572 TTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCA 1631
              |||||||||  |||| ||| || ||||||||  || ||||||||||||| ||||| ||||
Db           230 TTTGCTCATTTATAAAAAGCAACTCCTCATCCATTGAAGTTTTATCATGGGATTGTAGCA 289

Qy       1632 ATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTAC 1691
              || |||| |  |||||||||      | |||||||||| ||||||||||||||| ||
Db           290 ATTCAGTTACATCTTCAGGC------TTACTTTTAATTCTAGTTCTCTTGCTGTTTCCAC 343

Qy       1692 CACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGT 1751
              ||||| ||| ||  |||||| || |||| ||||||  | ||| ||||| ||||||||||
Db           344 CACATTTGTAGTGACTTCCTTCACTGAAATCTTGACTCCCTCAAAGTCGTCCATGAGGGT 403

Qy       1752 TGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATC 1808
              ||||||| |||||||||||| |||| || | ||| |||||||||    ||||||||||||
Db           404 TGGAATCAACTTCTTCCAAACTCCTATTGACATTGATATTTTGACTTCCTCCCATGAATG 463

Qy       1809 ATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACT 1868
              |  |  ||| |||||||||| || ||||||||||||||| || |||||||||| ||||||
Db           464 ACAATCGTTTTTAATGGCATCTAGAATGGTGAATCCTTTTCAGAAGGTTTTCAGTTTACT 523

Qy       1869 TAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAAT 1928
              |   ||||||| ||      |||||    |||| |  ||| ||| |||||  ||| | |||
Db           524 TTACCCAGATCTAT-----CAGAGTGATCACTGTTTATGACAGCTATAGATTTAAGAAAT 578

Qy       1929 GTAT------TTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTC 1982
              ||||      |||||  |||| ||| |||||||||  |  |||| ||||||||||  |   |
Db           579 GTATTTCGTATTCTTAGATAAGAAGACTTGAAAGTGG--ATTATTCCTTGATCCGTGGGC 636

Qy       1983 TGCAAAATAGATGTTGTG 2000
              || | ||| |||| ||||
Db           637 TGTAGAATGGATGGTGTG 654


RESULT 15
BX495939/c
LOCUS       BX495939                718 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp779E1425_r1 779 (synonym: hncc1) Homo sapiens cDNA clone
            DKFZp779E1425 5', mRNA sequence.
ACCESSION   BX495939
VERSION     BX495939.1  GI:32011089
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     EST (Bahr,A., Lauber,J., Mewes,H.W., Weil,B., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No s1 sequence available.
            This clone (DKFZp779E1425) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .718
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp779E1425"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="DH10B"
                     /clone_lib="779 (synonym: hncc1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB"
ORIGIN

  Query Match             16.5%;  Score 330;  DB 5;  Length 718;
  Best Local Similarity   74.1%;  Pred. No. 3e-42;
  Matches  551;  Conservative    0;  Mismatches  152;  Indels   41;  Gaps    9;
```

```
Qy        978 ACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATAT 1037
              ||| | ||||| || ||||| ||   |||||| || | |  ||||||||||
Db        718 ACACGTATACCATGAAGATATTGCAAGTTTGGGTCAAGATAACCACAATAAT-------- 667

Qy       1038 GCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATA 1097
                ||| |||||| |||| ||  || |||| || | |||||| |||  ||||||||
Db        666 -------GTGAATATCANAATATAGCCAGCCACATAA-TTTTTTGGTTTCTTAGTGCATA 615

Qy       1098 TAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTA-A 1156
              |||||||| || |||||||| ||||||||||||||||||| || ||  ||||||| | |
Db        614 TAAAAGTTATGTTTATACTATACTGTAGTCTGTTAAGTGTGTAACAGCATTATGTCGAGA 555

Qy       1157 AAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGA 1216
              |||    ||||||||| ||| || |  ||||||| ||| ||||||| ||||||||| |||
Db        554 AAATGTGCATACCTTAGTTTAAANACACTTTATTGCTACAAAATGCCAACAATCATCTGA 495

Qy       1217 GCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT- 1275
              || |||||||||||| |||||||||||| ||||| ||||||   ||    || |||  | |
Db        494 GCCTTCAGTGAGTT-TAATCTTTTTGCTAGTGGATGGTCTTGGCTCCATGTTGATGGCTG 436

Qy       1276 ----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA--A 1323
                       ||| |   | |||| ||| |  |||||||||||||| |||||||||  |
Db        435 CTAACTGATCAGGGTGGTGGTTGCTAAAGATGGGGGTGGCTGTGGCAATTTCTTAAAATA 376

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTG 1383
              |||| ||||  ||||   ||  ||||| |||||||||||||||||||||||||||| |||
Db        375 AAACAATGAAATTTGCTGCATTGATTGACTTTTCCTTTCATGAAAGATTTCTCTGTAGCA 316

Qy       1384 TGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAAT 1442
              ||  |   | | ||||| |||||||  |||||| |||| ||||| |||| |||| ||||
Db        315 TGCAACATTGTGTGATACCATTTTACCCACAGTGGAACCTCTTTCAAAACTGGAGCCAAT 256

Qy       1443 CCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGT 1502
              | |||||||||||||     |  |  | ||||||| |||| ||||||| ||||| ||||
Db        255 CTTCTCAAACCCTGCTAGTAATCTATCAACTAAGTTTATATGATATTCTAAATCCTTTGT 196

Qy       1503 TGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGA 1562
              |||||| |||||||| |||||||  |||||| || || |||||| ||||||||      |
Db        195 TGTCATCTCAACAATGTTCACAGCATCTTCATCAAGAATAGATTTCATCTCATGAAACCA 136

Qy       1563 GATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAG 1622
                |    |||||||   |||||||||||  || || |||||||||||||||||||| | ||
Db        135 CTT----TCTTTGCCTGTCCATAAGAAGCAACTCTTCATCTGTTCAAGTTTTATCCTAAG 80

Qy       1623 ATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGC 1682
              |||| || ||| |  |||| |||||||||     || ||||| |||||| ||||||||| |
Db         79 ATTGAAGTAATTCCTTCATATCTTCAGG-----CTCCACTTCTAATTCTAGTTCTCTTGG 25

Qy       1683 TGTTTCTACCACATCTGTGGTTCC 1706
              | ||||||||| ||||   ||| |
Db         24 TATTTCTACCATATCTACAGTTAC 1

RESULT 16
CR538654/c
LOCUS       CR538654                721 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  DKFZp459M217_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459M217 5', mRNA sequence.
ACCESSION   CR538654
VERSION     CR538654.1  GI:49684792
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pongo.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Deutschenbaur,S., Schaipp,A.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Ottenwaelder,B., Obermaier,B.,
            Deutschenbaur,S., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix
            (Martinsried/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp459M217) is available at
            the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum,
            Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email:
            clone@rzpd.de Further information about the clone and the
            sequencing project is available at
            http://mips.gsf.de/projects/cdna/.
FEATURES              Location/Qualifiers
     source           1. .721
                      /organism="Pongo pygmaeus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9600"
                      /clone="DKFZp459M217"
                      /tissue_type="cortex"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /clone_lib="459 (synonym: pcor1)"
                      /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             16.5%;  Score 329.8;  DB 7;  Length 721;
  Best Local Similarity   70.5%;  Pred. No. 3.2e-42;
  Matches  514;  Conservative    0;  Mismatches  199;  Indels   16;  Gaps    5;

Qy       1061 AGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACAC 1120
              ||  | |    |  | | |||||| |||||||||||||||||||  | |||  |  ||
Db        721 AGCAAATACAATGAATTTTTTGGTTTCCCAGTGCATATAAAAGTTACGTTTACNCCNCAG 662

Qy       1121 TGTAGTCTGTTAAGTGTGCAATAGTGTTATG-TCTAAAAAAACACATACCTTAATTTTAA 1179
              |||||| | |||||||||||||||  |||||| | | |||||  ||||||||||||||||
Db        661 TGTAGTGTATTAAGTGTGCAATAACGTTATGTTTTTAAAAAGTACATACCTTAATTTTAA 602

Qy       1180 AATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTT 1239
              ||| |||||||  |||||||  | |||||||||| |||||  |||||||||| |||||||
Db        601 AATACTTTATTGCTAAAAATAGTTAACAATCATCTGAGCCTTCAGTGAGTCATAATCTTG 542

Qy       1240 TTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAG 1299
              || | |  |    |  |  | ||    |          |    |||| |||    ||
Db        541 CTGGTAGAAGGTCTTGGCTCAATGTTATGGCTGCTGAATGATCAGGGTGATGACTGCTAA 482

Qy       1300 GGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCT 1359
              |     |  |   |   ||| |||||| | ||| |||| | ||||| |||||  |||||
Db        481 AGGTTGAGGTGTCTTAAAATAAGACAACAATAAAGTTTGCTACATCAGCTGACTTTTCCT 422

Qy       1360 TTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGA 1419
              |||| |||||| |||||| |||| ||   ||| | |||| ||||||||||   || | ||
Db        421 TTCACGAAAGAGTTCTCTGTAGTATGCAATGTTGTTTGTTAGCATTTTATCTGCAATTGA 362

Qy       1420 ACTTCTTTGAAAA--TTGGATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGT 1477
              || |||||  |||  | |  ||| || ||||||||| | |  |   || |  | |||| |
Db        361 ACGTCTTTTCAAAATTGGAGTCAGTCTTCTCAAACCATACCATTGCTTTATTAACTAAAT 302

Qy       1478 TAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAG 1537
              | || | |||||||||||||| |||||||||| || ||| | |||||||  |||| ||||
Db        301 TTATGTGATATTCTGAATCCTTTGTTGTCATCTCGACAGTGTTCACAGCATCTTTACCAA 242

Qy       1538 GAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCC 1597
              |||||||||| |||||        |  |     |||||| ||||||||| |||  |  | |
Db        241 GAGTAGATTCCTTCTCAAGAAACCACTT----TCTTTGGTCATCCATAGGAAACAGCTTC 186

Qy       1598 TCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACT 1657
              |||||    |||||| |||||||||||||||||||| ||| ||  ||||||      | ||
Db        185 TCATCATGTTAAGTTTGATCATGAGATTGCAGCAATTCAGCCACATCTTCA-----CTCT 131

Qy       1658 TCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTG 1717
                |||| |  ||| |||| ||||||| |||| ||||||||||  ||   |||||||  |
Db        130 CTACTTCTCCTTCTAGTTGTCTTGCTATTTCCACCACATCTGCAGTGATTTCCTCCACAG 71
```

```
Qy       1718 AAGTCTTGAACCTCTCCAAG----TCATCCATGAGGGTTGGAATCGACTTCTTCCAAATT 1773
              || ||||||  | ||| |||    || ||||||||||||| ||||  ||||| ||||| |
Db         70 AAATCTTGAGACCCTCAAAGGGGTTCGTCCATGAGGGTTGAAATCAGCTTCTCCCAAACT 11

Qy       1774 CCTGTTAAT 1782
              |||||||||
Db         10 CCTGTTAAT 2
```

RESULT 17
CK774525
LOCUS CK774525 831 bp mRNA linear EST 20-FEB-2004
DEFINITION 963293 MARC 2BOV Bos taurus cDNA 3', mRNA sequence.
ACCESSION CK774525
VERSION CK774525.1 GI:42728668
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 831)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
PUBMED 11282978
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross_match v0.990329.
Plate: 97 row: C column: 14
Seq primer: GTAATACGACTCACTATAGGG.
FEATURES Location/Qualifiers
source 1. .831
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2BOV"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI; Library made from pooled tissue from testis, thymus, semitendonosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."
ORIGIN

Query Match 16.4%; Score 328.2; DB 7; Length 831;
Best Local Similarity 71.2%; Pred. No. 5.6e-42;
Matches 606; Conservative 0; Mismatches 193; Indels 52; Gaps 11;

```
Qy       1179 AAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTT 1238
              ||||  |||||||||||||||  ||||||| ||||||| ||||||||||||||  ||||||
Db          1 AAATACTTTATTACTAAAACCTGCTAACCATCATTTGCGCATTCAGTGAGTCATAATCTC 60

Qy       1239 TTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CGGGGGTCAGGT 1287
              |||||| ||||| |||| |  || | ||   |  | |             || |  || |
Db         61 TTTGCTAGTGGAGGGTCCTGGCTCAGTGTACATGGCTGCTGACTGGTCAAGGTGGTAGTT 120

Qy       1288 GCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAA-----------CAGTGAAGAT 1336
              |||||||  | |||||||    |||| |||||||||| ||           | ||||| |
Db        121 GCTGAAGACTGGGGTGGTGAAGGCAATTTCTTAAAATAAGACAATAAGGCAAATGAAGTT 180

Qy       1337 TGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAG-----TGTGTGATGC 1391
              |||   ||||      || |||||| ||||||  |||  |||  |||         ||  ||||
Db        181 TGCGCCATCAACCTTCTTTTCCTTCCATGAACAATTATTCTGTAGCATGCATTGCAATGC 240

Qy       1392 TTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG-ATCAATCCTCTCAA 1450
              | |||  | ||| ||||| ||||| ||| ||||||| |||||||| ||||||||||||||
Db        241 TGTTTAGTGGCACTTTATCCACAGGAGAGCTTCTTTCAAAATTGGAATCAATCCTCTCAA 300

Qy       1451 ACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTT 1510
              |||||||       || |  | ||||||| || |||||||||| ||||| | || |||||||
Db        301 ACCCTGCCACTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTGGTCGTCATTT 360

Qy       1511 CAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAAT 1570
              |||| || |||| ||  |||||||||||||||||||||| |||||       ||| | |
Db        361 CAACCATCTTCATAGCATCTTCACCAGGAGTAGATTCCGTCTCA-------AGAAGCCAC 413

Qy       1571 CTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGC 1630
              ||  ||  ||       |||||  |||||| |  || ||||||||||||||||| |||||
Db        414 GTTCTCTGCTCAGCCGTAAGAAGCTCCTCAGCCATTAAAGTTTTATCATGAGATGGCAGC 473

Qy       1631 AATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTA 1690
              ||  || ||  || | ||||  ||||||     |||||| || | |||||||  ||| |
Db        474 AACTCACTCCCATCGTGAGGCTCCACTTC-----TAATTCTAGCTGTCTTGCTTCTTCCA 528

Qy       1691 CCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGG 1750
              |||||||| ||| ||||||||| ||||||||   |    || |||||| ||||||||
Db        529 TCACATCTGCAGTTACTTCCTCCACTGAAGTCTTGCCCTCTTCAAAGTCACCCATGAGGA 588

Qy       1751 TTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACC---TCCCATGAAT 1807
              ||||||| |||||||||| | ||| ||  || ||  ||||||||||   ||| ||||||
Db        589 CTGGAATCAACTTCTTCCACACTCCCGTGTATGTTGGTATTTTGACCTTTTCCTATGAAT 648

Qy       1808 CATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTAC 1867
              ||  |||||||||||||||  || |||||||||||||||||| | ||||||||||| ||
Db        649 CACAAATGTTCTTAATGGCATCTAGAATGGTGAATCCTTTCC-AGAGGTTTTCAATTGAC 707

Qy       1868 TTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAA 1927
              || | ||||||||||     ||| |||   | |  |  || ||| |||| ||| |  |||
Db        708 TTTGCCCAGATCCAT-----CAGTGGAATTATTACCTTTGGCAGCTATAACCTAAATGAA 762

Qy       1928 TGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAA-TTACTCCTTGATCCATTTTCTGCA 1986
              |  |||   |||||| |||||  |||||| |||  ||||||||||   |||||
Db        763 --ACATATTTCTTAAATAAGACTTGAGGGTTGAAATTTAGCCCTTGATCCATGGACTGCA 820

Qy       1987 AAATAGATGTT 1997
              |||  |||||
Db        821 GAATGTATGTT 831

RESULT 18
CR859436/c
LOCUS CR859436 5336 bp mRNA linear HTC 12-NOV-2004
DEFINITION Pongo pygmaeus mRNA; cDNA DKFZp469P156 (from clone DKFZp469P156).
ACCESSION CR859436
VERSION CR859436.1 GI:55729758
KEYWORDS HTC.
SOURCE Pongo pygmaeus (orangutan)
ORGANISM Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE 1 (bases 1 to 5336)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
CONSRTM The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Agowa (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp469P156) is available at the RZPD Deutsches
```

```
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp469P156
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES              Location/Qualifiers
     source           1. .5336
                      /organism="Pongo pygmaeus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9600"
                      /clone="DKFZp469P156"
                      /tissue_type="kidney"
                      /clone_lib="469 (synonym: pkid1). Vector pSport1_Sfi; host
                      DH10B; sites SfilA + SfilB"
                      /dev_stage="adult"
                      /note="mineralocorticoid receptor delta (Homo sapiens),
                      differentially spliced"
     gene             1. .5336
                      /gene="DKFZp469P156"
     CDS              93. .2036
                      /gene="DKFZp469P156"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="CAH91608.1"
                      /db_xref="GI:55729759"
                      /translation="MWLDLGGLRSRETKGYHSLPEGLGMERRWGQVSQAVEHSSLGPT
                      ERTDENNYTEIVNVSCVSGAIPNNSTQGSSKEKHELLPCLQQDNNRPGILTSDIKTEL
                      ESKELSATVAESMGLYMDSVRDADYSYEQQNQQGSMSPAKIYQNVEQLVKFYKGNGHR
                      PSTLSCVNRPLRSFMSDSGSSVNGGVMRAIVKSPIMCHEKSPSVCSPLNMTSSVCSPA
                      GINSVSSTTASFGSFPVHSPITQGTPLTCSPNVENRGSRSHSPAHASNVGSPLSSPLS
                      SMKSSISSPPSHCSVKSPVSSPNNVTLRSSVSSPANINNSRCSVSSPSNTNNRSTLSS
                      PAASTVGSICSPVNNAFSYTASGPSAGSSTSRDVVPSPDTQEKGAQEVPFPKTEEVES
                      AISNGVTGQLNIVQYIKPEPDGAFSSSCLGGNSKINSDSSFSVPIKQESTKHSCSGTS
                      FKGNPTVNPFPFMDGSYFSFMDDKDYYSLSGILGPPVPGFDGNCEGSGFPVGIKQEPD
                      DGSYYPEASIPSSAIVGVNSGGQSFHYRIGAQGTISLSRSARDQSFQHLSSPPPVNTL
                      VESWKSHGDLSSRRSDGYPVLEYIPENVSSSTLRSVSTGSSRPSKICLVCGDEASGCH
                      YGVVTCGSCKVFFKRAVEDVAVP"
ORIGIN

  Query Match           16.4%;  Score 327.4;  DB 4;  Length 5336;
  Best Local Similarity   67.1%;  Pred. No. 5.1e-42;
  Matches  715;  Conservative    0;  Mismatches  256;  Indels   94;  Gaps   14;

Qy        962 TAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACC 1021
              |||    | | |   ||||| ||| | || |||||| || | |||  |||||| | ||||
Db       4376 TAAGATCCAATAACATACAGACATGCTTTAGAGATATTGAGAGTTCTGTTCCAGATCACC 4317

Qy       1022 ACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTT 1081
              |||||||  |||| || | |  |  | |                  |||| |  | |||
Db       4316 ACAATAAAGCAAACATCCCAATAAAGCAAG----------------CACATACTTTTTTC 4273

Qy       1082 GGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAA 1141
              || |||||||  ||||| ||  ||   ||  |     |||| ||||| | || | ||||
Db       4272 GGTTTCCCAGCACATATTAAGTTTACACTATT-----ACTGAAGTCTATAAACTCTGCAG 4218

Qy       1142 TAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATG 1201
                |||  |   || |||||||   | ||       ||||||     |  |    |||||||
Db       4217 CAGTACTGCATCAAAAAAAATCTACACACCTTAATTAAAAATACCTTATTTCTAAAAATG 4158

Qy       1202 CTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC- 1260
              ||||  ||||  | ||| ||||||||||  |||||||| || ||| ||   |    ||
Db       4157 CTAATGATCA--TCAGCCTTCAGTGAGTAATAATCTTTTAGCGGGTAGAGTCTTGCCTCG 4100

Qy       1261 -------TTATTGATGACTGATCGGGGGTCAGGTGCTG--AAGCTTAGGGTGGCTGTGGC 1311
                    |  || |||||||| | |    |||      | |||| |||   |||||||
Db       4099 ATGTTGATGGCTGCTGACTGATCAGAGTGGTGGTTGCTGAAGGCTTGGGGCAACTGTGGC 4040

Qy       1312 AGTTTCTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGA 1366
              | |||||||||     |||||| ||||| ||||   | |  || ||| ||||||||||||
Db       4039 AATTTCTTAAAATAAGACAACAATGAAGTTTGCTGCAACGATTTACTTTTCCTTTCATGA 3980

Qy       1367 AAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTT 1426
              |||  |||||| |||  ||| ||||                          |||||||| |

Db       3979 AAGGGTTCTCTGTAGCATGTAATGC------------------------AGAACTTCCT 3945

Qy       1427 TGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAA 1485
              | |||| || |  |||||||||||| | || |      |  |  | |||||| |  |||
Db       3944 TCAAAACTGTAGGCAATCCTCTCAATCTCTACCACTGCTGTATCAACTAAGTTTACGTAA 3885

Qy       1486 TATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGAT 1545
              |||||| |||||  |||||||||||||||||| |||| ||  ||||| |||||||| ||||
Db       3884 TATTCTAAATCCTCTGTTGTCATTTCAACAATGTTCAAAGCATCTTCTCCAGGAGAAGAT 3825

Qy       1546 TCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGT 1605
              |||||||||       |  |    ||||| ||||||||| | ||  | ||||||||   |
Db       3824 TCCATCTCAAGAAACCACTT----TCTTTCCTCATCCATTAAAATCATTTCCTCATTAAT 3769

Qy       1606 TCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTT 1665
              |||||||||||| |||||||||||||||| ||||||  ||||||     |  | ||||| |
Db       3768 TCAAGTTTTATCCTGAGATTGCAGCAATTCAGTCACATCTTCA-----CGGTCCACTTCT 3714

Qy       1666 AATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTG 1725
              | ||| |||||||||| | ||||||||||||||||    || ||| |||| ||||| |||
Db       3713 ATTTCTAGTTCTCTTGTTATTTCTACCACATCTGCAATTACTTTCTCCAATGAAGTTTTG 3654

Qy       1726 AACCTCTCCAAGTCATCCATGAGG--------GTTGGAATCGACTTCTTCCAAATTCCTG 1777
                |||  ||| |||||||||||| |       ||||||||  ||||||||||| |||||
Db       3653 -ACCCCTCAAAGTCATCCATGATGGATGGAAAGTTGGAATTAACTTCTTCCAAACTCCTG 3595

Qy       1778 TTAATATTTATATTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACC------TG 1831
              ||||| || |||||||||||||||||||||| |  ||||| |  |  ||| |      |
Db       3594 TTAATGTTGATATTTTGACCTCCCATGAATCGTTTATGTTTTGGACAGCATCATATTATA 3535

Qy       1832 GAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGA 1891
              |||||||||| ||||||| ||  |||||||||||||| | |    ||       ||||
Db       3534 GAATGGTGAATTCTTTCCAGAAACTTTTCAATTTACTTTGCCAGATTC------ATCAGA 3481

Qy       1892 GGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTT 1951
              |||  |||| |  ||| |||| || |||||||  |||||||||||| |||||||| | |
Db       3480 GGAATCACTATGCATGGCAGTGATTGCCTTATAAAATGTATTTCTTAAATAATAAGACAT 3421

Qy       1952 GAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGT 1996
              ||||||  ||||| ||||| |||||||   |||| |||  ||||
Db       3420 GAAAGTCAAAATTCCTCCTGGATCCATGGACTGCAGAATGAATGT 3376


RESULT 19
BG476842/c
LOCUS       BG476842                929 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602525095F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4643519 5',
            mRNA sequence.
ACCESSION   BG476842
VERSION     BG476842.1  GI:13409121
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 929)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1413  row: m  column: 24
            High quality sequence stop: 738.
FEATURES              Location/Qualifiers
     source           1. .929
```

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4643519"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_20"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

  Query Match             16.3%;  Score 326.6;  DB 2;  Length 929;
  Best Local Similarity   78.3%;  Pred. No. 9.8e-42;
  Matches  474;  Conservative    0;  Mismatches  109;  Indels   22;  Gaps    6;

Qy       1386 TGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATC-- 1443
             |||| || |||||||||||||||  |||| ||||  |||||  |||| | |||| |||
Db        899 TGATTCTGTTTGATAGCATTTTACCCACATTAGACTTTCTTCAAAAAATTGATCCATCCC 840

Qy       1444 -CTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTG- 1501
              |||| |  |    | ||| || |   ||  |||||| | |||||| |||| || | |||
Db        839 TCTCTAATCCATGCCACTGGTTAATGTACTAAAGTTTAGATAATAATCTGGATTCCTTGA 780

Qy       1502 TTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTG 1561
              ||| ||||| ||||||| |||||||  ||||||| |||  |||||  ||     |  | |
Db        779 TTGGCATTGCAACAATGTTCACAGCTTCTTCACAAGGATCAGATTTCACTCAAGGTACCA 720

Qy       1562 AGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGA 1621
              |      |||||||||||||||||||||| |  ||||||||||||||||||  |||||||
Db        719 TG-----TTCTTTGCTCATCCATAAGAAGTATCTCCTCATCTGTTCAAGTTGGATCATGA 665

Qy       1622 GATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTG 1681
              |||||||||||| ||||||  ||||||||     || ||||| |||||| ||||||||||
Db        664 GATTGCAGCAATCCAGTCACATCTTCAGG-----CTCCACTTCTAATTCTAGTTCTCTTG 610

Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741
              ||  |||  |||||||||  |||||||||||||| |||||||||||||| ||| |||||||
Db        609 CTAGTTCCGCCACATCTGCAGTTCCTTCCTCCACTGAAGTCTTGAACCCCTCAAAGTCAT 550

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTG---ACCT 1798
              ||||  |||  ||||||| ||||||| |||| |||||||||| || | ||||||    |||
Db        549 CCATAAGGTATGGAATCCACTTCTTTCAAACTCCTGTTAATGTTGAAATTTTGGCTTCCT 490

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858
              |||||||||| |||||||||||||||| || || ||||||||| |||||| |||||||
Db        489 CCCATGAATCACAAATGTTCTTAATGGCATCTAGAGTGGTGAATCTTTTCCAGAAGGTTT 430

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
              | |||||||| |  |||||||||     |||||||  |||| || |||  || ||||||
Db        429 TTAATTTACTTTGCTCAGATCCAT-----CAGAGGAATCACTATCTATGGAAGATATAGC 375

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              |||||  |||||||||||||| ||||||| |||||||||||||||  ||||||||||||||
Db        374 TTTATGAAATGTATTTCTTCAGTAATAAGACTTGAAAGTTGAAATCACTCCTTGATCCAT 315

Qy       1979 TTTCT 1983
                ||
Db        314 GGGCT 310


RESULT 20
CE040574/c
LOCUS       CE040574                 755 bp    DNA     linear   GSS 24-SEP-2003
DEFINITION  tigr-gss-dog-17000350006572 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE040574
VERSION     CE040574.1  GI:35072379
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .755
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             16.3%;  Score 325.6;  DB 9;  Length 755;
  Best Local Similarity   72.7%;  Pred. No. 1.5e-41;
  Matches  528;  Conservative    0;  Mismatches  164;  Indels   34;  Gaps    7;

Qy       1277 GGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGAT 1336
              ||| | | | |||   || || || ||| |||||||  ||||||||   |||    | | |
Db        751 GGGTGGCGGTTGCGACAGGTTGGGTTGGGTGTGGCAATTTCTTAATGTAAC----ATGTT 696

Qy       1337 TGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTT 1396
              ||    ||   |||| ||||||||||||||    |||||||| | |  || | |||| |||
Db        695 TGGTGCATTGATTGAGTCTTCCTTTCATGATCATTTTCTCTGTGGCATGCGGTGCTGTTT 636

Qy       1397 GATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACCCTG 1456
              || |||||||||  |||||| | || ||||   || | | ||||||||||||||  |||
Db        635 GAAAGCATTTTACACACAGTGGCACGTCTTCAGAATTGGAATCAATCCTCTCAAAATCTG 576

Qy       1457 CTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAA 1516
                 || | | | ||    || || |||| | ||| | |||||| ||||||||||
Db        575 TAGACGCTTTATCAACCAACCATATGTAGTATTTTAAATTCTTTGTTGCCATTTCAACAA 516

Qy       1517 TTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGC 1576
              | |||||||   |||||||||||| | |||  ||| ||      || |     | ||||
Db        515 TCTTCACAGCACCTTCACCAGGAGGAAATTTTATCCCA-----AGAAACCATGTTATTGC 461

Qy       1577 TCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACA 1636
              ||||  |||||||| || || |||||| || ||| |||||||||| ||
Db        460 TCATTTATAAGAAGCAACTCTTCATCTCTTGAAGCTTTATCATGAGTTT----------- 412

Qy       1637 GTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACAT 1696
              ||||  |||||||||  |||||||     ||||| |||||||||||||| |||||  |||
Db        411 GTCACCTCTTCAGGCTCCACTTCA-----AATTCTAGTTCTCTTGCTGTCTCTACTGCAT 357

Qy       1697 CTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAA 1756
              |||| ||| ||||||||| || |||||||||||||| |||||| |||||| ||||||||
Db        356 CTGTAGTTACTTCCTCCACTG-AGTCTTGAACCTCTCAAAGTCAGCCATGATGGTTGGAA 298

Qy       1757 TCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACC---TCCCATGAATCATGAA 1813
              || |||| ||||||| ||||||| || || ||||||||||   |  ||||||||  |
Db        297 TCAACTTTTTCCAAACTCCTGTTTATGTTGATATTTTGACCACTTTTCATGAATCGCAAG 238

Qy       1814 TGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTC 1873
```

```
Db        237 TGTTCTTACCGACATCTAGAATAGTGGACTCTTTCCAGAAGGTTTTCACTTTGCTTTGTT 178

Qy       1874 CAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATT 1933

Db        177 CAGATCCAT-----CAGAAGAATCACTGTCTATGGAAGCTGTAGCCTTACTAAATGTATT 123

Qy       1934 TCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGA 1993

Db        122 TCTTAAATAATAATATGTGAATATTGAAATTACTCCTTGATCCATGGGCTGCAGAATGGA 63

Qy       1994 TGTTGT 1999

Db         62 TGTTGT 57


RESULT 21
AQ237450/c
LOCUS       AQ237450                815 bp    DNA     linear   GSS 21-APR-1999
DEFINITION  RPCI11-61M13.TJB.1 RPCI-11 Homo sapiens genomic clone
            RPCI-11-61M13, genomic survey sequence.
ACCESSION   AQ237450
VERSION     AQ237450.1  GI:3669741
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 815)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .815
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7523340"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-61M13"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             16.3%;  Score 325.6;  DB 9;  Length 815;
  Best Local Similarity   73.3%;  Pred. No. 1.4e-41;
  Matches  595;  Conservative    0;  Mismatches  169;  Indels   48;  Gaps   12;

Qy       1049 ATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTG 1108

Db        783 ATACCACAATAAAGTGAGCCACAGAA-TTTTTCAGTTTCCCAGTGCCTGCAAAAGTTATG 725

Qy       1109 CTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAA-----CA 1163

Db         724 TTTATAACATACTGCAGTCTATTAAATGTGCATTAGCATTGTGTCTAAAAATACAATGCA 665

Qy       1164 CATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCA 1223

Db         664 CATAACGTTAATTTAAAGATACTTTATTGCTTAATATGCTAACAATCATCTGAGCCTTCA 605

Qy       1224 GTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT---------TTTCTTATTGATGACTGA 1274

Db         604 GCAAATCATAATCTTTTTTTGGTGGAAAGTTTTGCCTCGGTATTGGTGGCTGATGACTCA 545

Qy       1275 TC-GGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAA 1333

Db         544 TCAGGGTGGTGCTTGCTGAAGGTTTGGGTGGCTGTGGCAATTTCTTAAAATAAC----AA 489

Qy       1334 GATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTT 1393

Db         488 GTTTGCCCCATCAATTGACTCTTCCTCTCGTGAAAGACTTCTCTGTAGCGTGCGAAGCTG 429

Qy       1394 TTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACC 1453

Db         428 TTTTATAGCATTTTGCCCAGAGTAGAAGTTCTTTCAAAA----GTCAATTCTCTCAAACT 373

Qy       1454 CTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAA 1513

Db         372 CTGCTACTGCTTTA----CCAAGTATATGTAATATTTTAAATCCTT---CATCATTTCAG 320

Qy       1514 CAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTT 1573

Db         319 CAATACTCACAGCATCCTCACCAGAAGTAAACTCCATCTCAGAT----AACTACGTTCTT 264

Qy       1574 TGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAAT 1633

Db         263 TGCTCATCCATGAGAAACAACAACTCATCTTTTCAAATTTTATCATGAGATTGCAGCAAT 204

Qy       1634 ACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCA 1693

Db         203 GTAGTCTCATCTTCAGG-----ATTCATTTCTAATTACAGTTCTTTTGTTATTTCCACCA 149

Qy       1694 CATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTG 1753

Db         148 TATCTG-----CCTTTCTCCATTGAAATCTTGAATCCCTCAAAGTCATCCATGAAGGTTG 94

Qy       1754 GAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC---CTCCCATGAATCAT 1810

Db          93 GAGTAAACTTCTTTCAAGCTTTTGTTAATGTTGATATTTTGACTTACTCCCTTGAATCAT 34

Qy       1811 GAATGTTCTTAATGGCACCTGGAATGGTGAAT 1842

Db          33 GAATGTTCTTAATAACATCTAGAATGGTGAAT 2


RESULT 22
BU622127
LOCUS       BU622127                705 bp    mRNA    linear   EST 23-SEP-2002
DEFINITION  UI-H-FH1-bfu-m-19-0-UI.s1 NCI_CGAP_FH1 Homo sapiens cDNA clone
            UI-H-FH1-bfu-m-19-0-UI 3', mRNA sequence.
ACCESSION   BU622127
VERSION     BU622127.1  GI:23288342
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Martin
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
```

```
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be obtained
           from Dr. M. Bento Soares, bento-soares@uiowa.edu
            The following repetitive elements were found in this cDNA
           sequence: 1-38, >AT_rich#Low_complexity 39-705,
           >TIGGER1#DNA/MER2_type (matched compliment)
           Seq primer: M13 FORWARD
           POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FH1-bfu-m-19-0-UI"
                     /tissue_type="Cell Line"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_FH1"
                     /note="Organ: Chondrosarcoma; Vector: pT7T3-Pac
                     (Pharmacia) with a modified polylinker; Site_1: EcoR I;
                     Site_2: Not I; NCI_CGAP_FH1 is a normalized cDNA library
                     obtained from a cell line derived from grade I
                     chondrosarcoma tissue. The library was constructed and
                     normalized according to Bonaldo, Lennon and Soares, Genome
                     Research, 6:791-806, 1996. First strand cDNA synthesis was
                     primed with an oligo-dT primer containing a Not I site.
                     Double stranded cDNA was ligated to an EcoR I adaptor,
                     digested with Not I, and cloned directionally into
                     pT7T3-Pac vector. The oligonucleotide used to prime the
                     synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AGAATCCGGC. The cell line was provided by Dr. James Martin
                     from the University of Iowa.
                     TAG_TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
                     Chondrosarcoma
                     TAG_LIB=UI-H-FH1
                     TAG_SEQ=AGAATCCGGC"
ORIGIN

  Query Match             16.3%;  Score 325.4;  DB 5;  Length 705;
  Best Local Similarity   75.4%;  Pred. No. 1.6e-41;
  Matches  522;  Conservative    0;  Mismatches  132;  Indels   38;  Gaps    8;

Qy        974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
              || |||||||||||||   ||||| ||||| |||||||||| |||||    ||||
Db         34 TAATACAGGCATACCTCAAAGATATTGTGGATTTGGTTCCAGACCACTGTAATA------ 87

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
                  ||| | |||||| |||||||||||  ||||| | |||||  |||||||||
Db         88 ---------AAGCGAATATCATAATAAAGTGAGTTTCACAACTTTTTTGT-TTCCCAGTG 137

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
                |||||||||| || ||| |||| |||| ||||| ||||| || |||||   |||||||
Db        138 AGTATAAAAGTTATGTTTACACTATACTGCAGTCTATTAAGGGTTCAATAACATTATGTC 197

Qy       1154 TAAAAAA-ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212
              |||||||    | | |||||||||| ||||| |||||||||||||||| ||||  |||||
Db        198 TAAAAAATGTATACACCTTAATTTAAAAATACTTTATTACTAAAAAATACTAATGATCAT 257

Qy       1213 TTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT----------TTTCTT 1262
                | ||| ||||||| || || |||||||||||||||| |||||          ||  |
Db        258 CTAAGCCTTCAGTGGGTCGTCATCTTTTTGCTGGTGGAGGGTCTTGCCTCGATGTTGATG 317

Qy       1263 ATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA 1322
                || ||| ||||| | ||   ||      |      || |||||||||| |||||||||
Db        318 GCTGCTGATTGATCAGAGGAGTGGATGCTGAAGGCTAGGGGGCTGTGGCAATTTCTTAAA 377

Qy       1323 -----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCT 1377
                  |||||| ||||| |||||   |||| ||||||||||||||||||||||||||||||
Db        378 ATAAGACAACAATGAAGTTTGCTGCATCAATTGACTCTTCCTTTCATGAAAGATTTCTCT 437

Qy       1378 CTAGTGTGTGATGCTTTTTGATAGCATTTTA-TGCACAGTAGAACTTCTTTGAAAATTGG 1436
              ||||||||||||||| |||||||||||||||   || |||||||||||||| |||||| |
Db        438 GTAGTGTGTGATGCTGTTTGATAGCATTTTACCCCATAGTAGAACTTCTTTCAAAATTTG 497

Qy       1437 A-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAAT 1495
              | |||| |||||||||| |||      ||| |  | ||||||| || ||||||||  |||
Db        498 AGTCAAGCCTCTCAAACCATGCCACTGTTTTATCAACTAAGTTTATGTAATATTCAAAAT 557

Qy       1496 CCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCAT 1555
              || |||||| |||||||||| |  ||||| |||||||||||||||||| |||||| |||
Db        558 CCTTTGTTGCCATTTCAACAGTGATCACACTGTCTTCACCAGGAGTAGGTTCCATATCAA 617

Qy       1556 TTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTA 1615
                 |  |    |||||| ||||| |||||||  || ||||||||  |||||||| |
Db        618 GAAACCACTT----TCTTTGTTCATCAATAAGAAACAACTCCTCATCCATTCAAGTTTGA 673

Qy       1616 TCATGAGATTGCAGCAATACAGTCATGTCTTC 1647
              ||||  |||||||||||| |||||   |||||
Db        674 TCATCGGATTGCAGCAATTCAGTCNCATCTTC 705


RESULT 23
AQ748474
LOCUS       AQ748474                844 bp    DNA     linear   GSS 19-JUL-1999
DEFINITION  HS_5540_A2_E04_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1116 Col=8 Row=I, genomic survey sequence.
ACCESSION   AQ748474
VERSION     AQ748474.1  GI:5535632
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1116  row: I  column: 8
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 844.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate:1116 Col=8 Row=I"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN
```

```
  Query Match             16.3%;  Score 325.2;  DB 9;  Length 844;
  Best Local Similarity   75.9%;  Pred. No. 1.7e-41;
  Matches  486;  Conservative    0;  Mismatches  133;  Indels   21;  Gaps    6;

Qy       1369 GATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTG 1428
             || |||||| |||    |  ||||| |  ||||||||||||  ||||||||||||||||||
Db         15 GAATTCTCTGTAGCTAGAAATGCTGTACGATAGCATTTTACCCACAGTAGAACTTCTTTC 74

Qy       1429 AAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATA 1487
             |||||||||  |||||||  ||||||| ||||     || |   | ||||||| | |||||
Db         75 AAAATTGGAGACAATCCTTTCAAACCTTGCTGCTGCTTTATCAACTAAGTTACTGTAATA 134

Qy       1488 TTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTC 1547
             |||| ||||  |||| | |||||||  | | |||||||  |||  ||||| |||||||||
Db        135 TTCTAAATCTTTTGTGGCCATTTCATTAGTGTTCACAGCATCTCTACCAGTAGTAGATTC 194

Qy       1548 CATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTC 1607
               |||||       |  |    | ||| ||||||||||| | | |  |  |||| ||||
Db        195 TGTCTCAAGAGGCCACTT---TTTTTTTCTCATCCATAAAAGGTATCTGTTCATATGTTA 251

Qy       1608 AAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAA 1667
             |||||||| |||||||| |||| || ||||||  |||||||      |||||||| | |
Db        252 AAGTTTTACAATGAGATTTCAGCTATTCAGTCACATCTTCAG-----CCTTCACTTCTCA 306

Qy       1668 TTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAG----TCT 1723
             |||   |||||||||  |||| ||||||||||  ||| ||||||||| |||||    |||
Db        307 TTCTTATTCTCTTGCCTTTTCCACCACATCTGCAGTTACTTCCTCCACTGAAGTCTTTCT 366

Qy       1724 TGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATA 1783
             |||||| ||  ||||||||| |||||| ||||||| |||||||| |||| ||| |||||
Db        367 TGAACCCCTTGAAGTCATCCTTGAGGGGTGGAATCAACTTCTTCTAAATGCCTCTTAATC 426

Qy       1784 TTTATATTTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGA 1840
             || |||||||||   |||||||  ||||||||||||||||||||| || ||| |||||
Db        427 TTGATATTTTGACCTCCTCCCACAAATCATGAATGTTCTTAATGGCATCTAGAAGGGTGA 486

Qy       1841 ATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACT 1900
               |||||||| |||||||||||||||||| |  |||| |||    |||||||  |||
Db        487 GTCCTTTCCAGAAGGTTTTCAATTTACTTTGGTCAGACTCAT-----CAGAGGAATCACC 541

Qy       1901 TTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGA 1960
               || ||| ||| ||| || |||   |||||||||||| | |||||||  ||||||| ||||
Db        542 ATCTATGGCAGCTATTGCTTTACAAAATGTATTTCTTGAGTAATAAGACTTGAAAATTGA 601

Qy       1961 AATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
             |||||| |||||| ||||    ||||| | |   || |||||
Db        602 AATTACCCCTTGACCCATGGGCTGCAGATTGAATATTGTG 641


RESULT 24
CC478317/c
LOCUS       CC478317                 849 bp    DNA     linear   GSS 16-JUN-2003
DEFINITION  CH240_304N14.T7 CHORI-240 Bos taurus genomic clone CH240_304N14,
            genomic survey sequence.
ACCESSION   CC478317
VERSION     CC478317.1  GI:31755580
KEYWORDS    GSS.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 849)
  AUTHORS   Holt,R., Stott,J., Yang,G., Barber,S., Smailus,D., Prabhu,A.-L.,
            Tsai,M., Cloutier,A., Lee,D., Girn,N., Olson,T., Mayo,M.,
            Butterfield,Y., Kirkpatrick,R., Liu,J., Guin,R., Chan,A., Chiu,R.,
            Mathewson,C., Wye,N., Masson,A., Brown-John,M., Jones,S.,
            Schein,J., Marra,M., de Jong,P., McWilliam,S., Barris,W.,
            Dalrymple,B.P. and Tellam,R.
  TITLE     Bovine BAC End Sequences from Library CHORI-240, PLATES 294 to 398
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CH240_304N14.TARBAC13P2
            Contact: Rob Holt
            Sequencing
            The British Columbia Cancer Agency Genome Science Centre
            600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
            Tel: 604-877-6085
            Fax: 604-877-6276
            Email: rholt@bcgsc.ca
            Clones are derived from the bovine BAC library CHORI-240
            (http://www.chori.org/bacpac/bovine240.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/ordering_information.htm). This work
            was undertaken as part of the International Bovine BAC Mapping
            Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the
            British Columbia Genome Sciences Centre, Canada.
            Plate: 304  row: N  column: 14
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .849
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="breed: Hereford"
                     /db_xref="taxon:9913"
                     /clone="CH240_304N14"
                     /sex="Male"
                     /cell_type="Blood"
                     /clone_lib="CHORI-240"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
                     library (Male) produced by Pieter de Jong"
ORIGIN

  Query Match             16.2%;  Score 325;  DB 9;  Length 849;
  Best Local Similarity   70.1%;  Pred. No. 1.8e-41;
  Matches  581;  Conservative    0;  Mismatches  205;  Indels   43;  Gaps    9;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
               ||  | ||| || ||||||||||||||||  | | |||||| ||||||||||| |||||
Db        811 AAACGAATACCATAATAAAGTGAGTCACATGAATTTTTTGGTTTCCCAGTGCATGTAAAA 752

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC-TAAAAAAA 1161
              ||| || ||| |||| |||||||||| |||||||  |||  |  ||||||| | |||||
Db        751 GTTATGTTTACACTATACTGTAGTCTATTAAGTGCACAACTGAATTATGTCTTTAAAAAG 692

Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATT 1221
              |||| |||||||||  |||||  || |||  |||||||||||||||  |||| ||||| ||
Db        691 TACATTCCTTAATTTAAAAATATTTCATTGTTAAAAAATGCTAACTGTCATCTGAGCCTT 632

Qy       1222 CAGTGAGTTGTAA---TCTTTTTGCTGGTGGAAGGTCTTTTCTT----------ATTGAT 1268
              |   |||||| |   | |||||  |||||| ||||||  |||           || |
Db        631 TAACAAGTTGTCAGTCTTTTTTTTTTGGTGGAGGGTCTTGCCTTGATGTTGATAGCTGCT 572

Qy       1269 GACTGATCGGGGGTCAGGTGCTGAAGCTTAGG----GTGGCTGTGGCAGTTTCTTAAAAC 1324
              |||||||| | | |   || |   |  ||    || ||||||||| ||||||||||
Db        571 GACTGATCAGTGTTGTAGTTCCTCAAAGTCGGGGTAGTTGCTGTGGCAATTTCTTAAAAT 512

Qy       1325 AACAG---TGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAG 1381
              || ||   ||||| ||||   ||   | |||||||||||||| ||   |||||| | |||
Db        511 AAGAGTAATGAAGTTTGCTCCATTGATCGACTCTTCCTTTCAGGAGCAATTTCTTTGTAG 452

Qy       1382 TGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAA 1441
               ||   |||| ||||| |||||||  |  ||||     |  |  ||  ||| | |  |||
Db        451 CATGCAGTGCTGTTTGACAGCATTTAACCCACATAGAACTTCTTTCAAAATTAGAGTCAG 392

Qy       1442 TCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTG 1501
              |||||||||| |||| |      ||     | || || | || ||||||||| ||||| |||
Db        391 TCCTCTCAAATCCTGATGCTGCTTTCTCAACTTAGCTGATTTAATATTCTAAATCCTTTG 332

Qy       1502 TTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTG 1561
              |||||||||||||||  ||||||   |||||||||| |||||||| | ||||
Db        331 TTGTCATTTCAACAACCTTCACAACATCTTCACCAGAAGTAGATTCTAGCTCAAGAAACC 272
```

```
Qy       1562 AGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGA 1621
Db        271 ACTT----TTTTTGCTCACCCATAGGAAGCAACGTTTCATCCATTAAATCTGTATCACGA 216

Qy       1622 GATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTG 1681
Db        215 TATCACAGCAATTCAGACATATCTCCAGGCTCCAC----------TTTCAGTTATCTTG 167

Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741
Db        166 CTATTTCCGTCACATCTGCGGGTACAGCTGCCACCGAAGCCTTGGACCCCTCAAAGTCAT 107

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT--- 1798
Db        106 CCATGAAGGCTGGAGTGAA---CTTCCAAACTCCTGTTCATGTTGATATTTTGGCCTCTT 50

Qy       1799 -CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTT 1846
Db          49 CCCCATGAATCCTGAATGTTCTTAATGACGTCTAGGATAGTGAGTTCAT 1
```

RESULT 25
BQ067736/c
LOCUS BQ067736 1039 bp mRNA linear EST 02-APR-2002
DEFINITION AGENCOURT_6643057 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5768547 5', mRNA sequence.
ACCESSION BQ067736
VERSION BQ067736.1 GI:19896782
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1039)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12828 row: j column: 04
High quality sequence stop: 655.
FEATURES Location/Qualifiers
source 1. .1039
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5768547"
/lab_host="DH10B"
/clone_lib="NIH_MGC_121"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."
ORIGIN

Query Match 16.2%; Score 324.2; DB 3; Length 1039;
Best Local Similarity 70.4%; Pred. No. 2.3e-41;
Matches 589; Conservative 0; Mismatches 208; Indels 40; Gaps 10;

```
Qy       1059 AAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTAC 1118
Db        829 AAGGCAAATCACACAAATCTTTTGTTTGCTGGTGCAAATAGAACTGCGTGTATACAATAT 770

Qy       1119 ACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTA 1178
Db        769 ATAGTAGTCTAATAAGTGTACGA-CGGCATATGTCTAAAAATGTATATACCTTAATTTAG 711

Qy       1179 AAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAA---- 1234
Db        710 AATATTTTATTCATACAAATTATTTAATAATCATGTGAGCCTTCAATGAGTCTTTTGTTG 651

Qy       1235 ----------TCTTTTTGCTGGTGGAAGGTCTTTTCTTAT-----TGATGACTGATCGGG 1279
Db        650 TTGTAGTTGTTGTTGTTGTTGATGGAGTGTGTCAACATAGATGGCTGCTGACTGATCA-G 592

Qy       1280 GGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAG 1334
Db        591 GGTGATGTGCTGAAGGTAGGGGTGGCTGTGGCAATTTCTTAAAGTAACACAAAAATGATG 532

Qy       1335 ATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTT 1394
Db        531 TTTATCAGATCAATTGACTCTTCCTTTTAGGAAAGATTTGTCTGTGGCACACAATACTGT 472

Qy       1395 TTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACCC 1454
Db        471 TTGATAGCATTTTACCCATGGAAGAACT--TTCAAAATTGGAATCAATCTTTGCAAATCC 414

Qy       1455 TGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAAC 1514
Db        413 TGCCTCTGCTTTATCAACTACATT-ATGGACTATTCTTAATCCTTTGTTGTCATTTTAAC 355

Qy       1515 AATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTT 1574
Db        354 AATGTTCACAGCGTCTTCACCAGGAATAGATTCTGTCTCAAGAAACCACTT----TCTTT 299

Qy       1575 GCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATA 1634
Db        298 GCTCATCCATAAGAAGCAATTCCTCATGTGTTCAAGTTTTAGAATGAGATTGCAGCAATT 239

Qy       1635 CAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCAC 1694
Db        238 CCTTGACATCTTCAGG-----CTACACTGTCAATTCTAGTTCTTTTGCTATTTCTACCAC 184

Qy       1695 ATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGG 1754
Db        183 ATCTGCTGTTAATCCCTCTACTGAAGTTTTGAAGCCCTCAAAGTCATCCAAGAGGGTTGT 124

Qy       1755 AATCGACTTCTTCCAAATTCCTGTTAATATTTATAT--TTTGACCTCCCATGAATCATGA 1812
Db        123 GATCAACTTCATCCAAACTCCTGAAAATATCAATATGAATCTATCTCCCATGAATCACTA 64

Qy       1813 ATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTT 1869
Db          63 ATGTTCTTAATGGTGTCTAGAATGGTGAATATTTTCCAGGAGGTTTTCAATTTACAT 7
```

RESULT 26
AQ750123/c
LOCUS AQ750123 853 bp DNA linear GSS 19-JUL-1999
DEFINITION HS_5575_A2_C03_SP6 RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1151 Col=6 Row=E, genomic survey sequence.
ACCESSION AQ750123
VERSION AQ750123.1 GI:5537281
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 853)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

```
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1151  row: E  column: 6
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 853.
FEATURES             Location/Qualifiers
     source          1. .853
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=1151 Col=6 Row=E"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             16.1%;  Score 321.4;  DB 9;  Length 853;
  Best Local Similarity   73.8%;  Pred. No. 6.6e-41;
  Matches  538;  Conservative    0;  Mismatches  136;  Indels   55;  Gaps    8;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
             ||||| ||||| |||||||| ||| |||||  | |||    || | ||||||||||||||
Db        691 AAGTGAATATCGCAATAAAGCAAGTAACACACATTTTT----TTTCTAGTGCATATAAAA 636

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC 1162
             |||  | ||| |||    ||||||||| |||||||||
Db        635 GTTAGGTTTACACTCTTCTGTAGTCTATTAAGTGTGT----------------------- 599

Qy       1163 ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTC 1222
             ||||||||| |||   |||   ||    |  |||||||||  |||||| ||||   ||
Db        598 ACATACCTTCATT---AAAATAATTTACTCCTAAAAATGCTAGTAATCATCTGAGTCATC 542

Qy       1223 AGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTA----------TTGATGACT 1272
             |||||||| ||||||| || ||||| || ||||||  || |           || || ||
Db        541 AGTGAGTTTTAATCTTCTTTCTGGTAGAGGGTCTTGCCTCAGTGTTGGTGGCTGCTGTCT 482

Qy       1273 GATCGGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAAC 1327
             |||||||||    |  ||||||| ||   |||||||| | || |||||||||     |||||
Db        481 GATCGGGGGGGTGGTAGCTGAAGATTGTGGTGGCTGCGACAATTTCTTAAAATAAGACAAC 422

Qy       1328 AGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTG 1387
             | ||||||||||   |||   |||||||||||||  |  ||||||||||| |||  ||||
Db        421 AATGAAGATTGCTGCATCGACTGACTCTTCCTTTGACAAAAGATTTCTCTGTAGCATGTG 362

Qy       1388 ATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTC 1446
             |||||  || ||||||||||||   || ||||||||||||| |||||||| ||||||||
Db        361 ATGCTGTTGGATAGCATTTTACCCAGAGTAGAACTTCTTTCAAAATTGGAGTCAATCCTT 302

Qy       1447 TCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTC 1506
             ||| || |||||     || |  | ||||||| || || |||||| ||||| |||||||||
Db        301 TCATACTCTGCTGCTGCTTCATCAACTAAGTTTATGTACTATTCTAAATCCTTTGTTGTC 242

Qy       1507 ATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATG 1566
             ||||||||||  ||||||  ||||||| |||||||| ||||||||||        | |
Db        241 ATTTCAACAATGCTCACAGAATCTTCACTAGGAGTAGCTTCCATCTCAAGAGACCACTT- 183

Qy       1567 GAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTG 1626
             ||||||||  | || ||||||| || ||||||| |||||||||| |||||||| |||
Db        182 ---TCTTTGCTAACCCGTAAGAAGCAACTCCTCATATGTTCAAGTTTGATCATGAGTTTG 126

Qy       1627 CAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTT 1686
             |||||||  ||||||  | ||||||      || ||||| |||||| | ||||||| || ||
Db        125 CAGCAATTCAGTCACATGTTCAGG-----CTCCACTTCTAATTCTATTTCTCTTACTTTT 71

Qy       1687 TCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATG 1746
             || ||||||||||  ||| |||||||| |||||||||| || | ||| ||||||||||||
Db          70 TCCACCACATCTGCTGTTACTTCCTCCACTGAAGTCTTAAAGCCCTCAAAGTCATCCATG 11

Qy       1747 AGGGTTGGA 1755
             |||||||||
Db          10 AGGGTTGGA 2


RESULT 27
BG028427/c
LOCUS       BG028427                931 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  602294332F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4389342 5',
            mRNA sequence.
ACCESSION   BG028427
VERSION     BG028427.1  GI:12417521
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 931)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10076  row: o  column: 07
            High quality sequence stop: 681.
FEATURES             Location/Qualifiers
     source          1. .931
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4389342"
                     /tissue_type="osteosarcoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_86"
                     /note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.533 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             15.9%;  Score 318;  DB 2;  Length 931;
  Best Local Similarity   74.6%;  Pred. No. 2.2e-40;
  Matches  491;  Conservative    0;  Mismatches  135;  Indels   32;  Gaps    6;

Qy       1360 TTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGA 1419
             | | |  | ||    ||||||  |  ||     | |||||| |||||||  || ||||||
Db        697 TGCCTTCAGGAGGATTCTCTATGGCATGTGATCTGTTGATAACATTTTACCCAGAGTAGA 638
```

```
Qy       1420 ACTT--CTTTGAAAATTGGATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACC----- 1472
Db        637 ACTTTCTTTCAAAATTGGAGTCAACCTTCTCAAACCCTCCGCTGCTTTATCAACTAAGTA 578

Qy       1473 --------TAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACA 1524
Db        577 ATGTAATGTAAAGTTATGTAATATTCTCAGTCCTTTGTTGTCATTTCAACAATATTCACA 518

Qy       1525 GTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCAT 1584
Db        517 GCGTCTTCAACAGGAGTAGATTCTATCTCAAGAAACCACTT----TCTCTGCTCATCTGT 462

Qy       1585 AAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTC 1644
Db        461 GAGAAGCAACTACTCACCCAATCAAGTTTGATCAATAAATTGCAGCAATTTGGTTGCATC 402

Qy       1645 TTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTT 1704
Db        401 TCCAGG-----CTCCACTTCTAATGCTAGTTCTCTTGCTATTTCTACCACATCTGCAGTT 347

Qy       1705 CCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTC 1764
Db        346 CCCTCCTCCAATGAAGTCTTGAACCCTTCAAAGTGATCCATGAGGGCTGGAATCAGTTTC 287

Qy       1765 TTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATCATGAATGTTCTTA 1821
Db        286 TTCCAAACTCCTGTTAATGTTGATATTTTGACCTCCTCCCATGAATCATGAATATTCTTA 227

Qy       1822 ATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCA 1881
Db        226 ATAGCATCTAGGATGGTGACTCCTTTCCAGAAGGTTTTCAATTGACTTTGCCCAGATCCA 167

Qy       1882 TCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAAT 1941
Db        166 T-----CAGAGGAATCACTATCTATGGCAGCTGTAGCTTTATGAAATGTATTTATTAAGT 112

Qy       1942 AATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGT 1999
Db        111 AATAAGACTTGAAAGTCAAAATTACTCCTTGACTTCTGGGTTGTAGAATGGATGTTGT 54
```

RESULT 28
CB853133

```
LOCUS       CB853133                693 bp    mRNA    linear   EST 22-APR-2003
DEFINITION  UI-CF-FN0-afq-e-10-0-UI.s1 UI-CF-FN0 Homo sapiens cDNA clone
            UI-CF-FN0-afq-e-10-0-UI 3', mRNA sequence.
ACCESSION   CB853133
VERSION     CB853133.1  GI:30048271
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com) or from Open Biosystems
            (www.openbiosystems.com).
            The following repetitive elements were found in this cDNA
            sequence: 19-693, >TIGGER1#DNA/MER2_type (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-FN0-afq-e-10-0-UI"
                     /tissue_type="Human Lung Epithelial cells"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-CF-FN0"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-FN0 is a subtracted cDNA library derived from two
                     normalized Human lung epithelial cell libraries (EN1 and
                     DU1) The library was subtracted according to according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. For additional information, contact:
                     bento-soares@uiowa.edu
                     TAG_SEQ=None found"
ORIGIN
```

Query Match 15.8%; Score 316.4; DB 6; Length 693;
Best Local Similarity 73.6%; Pred. No. 4.2e-40;
Matches 509; Conservative 0; Mismatches 151; Indels 32; Gaps 7;

```
Qy       1058 TAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTA 1117
Db         17 TTAAGGGAGTCACACAATTATTTTTATTTTCCAGTACCTATAAAAGTTATGTTTAATCTA 76

Qy       1118 CACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTT 1177
Db         77 TATTATAGTCTATTAAGTTTGCAATAGCATTATATCTGACAATGTACATACTTTAATTTA 136

Qy       1178 AAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCT 1237
Db        137 AAAATATTTTATTGCTAAAAGATGCTAACGATAATCTAAGTCTTCAGTGAGT-----CTT 191

Qy       1238 TTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CGGGGGTCAGG 1286
Db        192 TGTTGCTGGAAGACAGTCTTGCCTCACTGTTGATGGCTGCTGAATGATCGGTGTGGTGGT 251

Qy       1287 TGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGCAA 1341
Db        252 TGCTGAAGGTTAGTGTGGCTCTAGAGATTTCTTTAAATGAGACAACAAGGAAGGCTGCAA 311

Qy       1342 TATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAG 1401
Db        312 CATCAACTGACTCTTCCTTTC-CGAAAGATTTATCTGTAGCATGTGTTGCTGTTGGATAG 370

Qy       1402 CATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCT 1460
Db        371 CCTTTTACCCACAGTGGAACATTCTTCAAAATTGAAGTCAATCCTCTCAAACTCTGCTGC 430

Qy       1461 GCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTT 1520
Db        431 TGTTTTATCAACTAAGTTTATGTAATATTCTAAGTCCTTTGTTGTCATTTCAACAATGGT 490

Qy       1521 CACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCAT 1580
Db        491 CACAGGATCTTCACCAGGAGTAGTTTCCATCTCAAGAAACCACTT----TTGTGGCTCTT 546

Qy       1581 CCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCA 1640
Db        547 ATATTAGAAGCAACTCCTCATTTGTTCACGTTTTATCATGAGATTGCAGCAATTCAGTCA 606

Qy       1641 TGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGT 1700
Db        607 CATATGCAGG-----TTCCACTTCTAAGTCTAGTTATCTTGCTATTTCCACCACATCTGC 661
```

```
Qy       1701 GGTTCCTTCCTCCATTGAAGTCTTGAACCTCT 1732
              ||| |||| |||| ||||||||||||||| ||
Db        662 AGTTACTTCGTCCACTGAAGTCTTGAACCCCT 693


RESULT 29
CF126937/c
LOCUS       CF126937                703 bp    mRNA    linear   EST 05-AUG-2003
DEFINITION  UI-HF-ET0-avx-m-05-0-UI.r1 NIH_MGC_214 Homo sapiens cDNA clone
            IMAGE:30563524 5', mRNA sequence.
ACCESSION   CF126937
VERSION     CF126937.1  GI:33204674
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Mary Hendrix
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/humanfl.html
             The following repetitive elements were found in this cDNA
            sequence: 1-701, >TIGGER1#DNA/MER2_type
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .703
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30563524"
                     /tissue_type="Chondrosarcoma Lung Metastasis cell lines"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_MGC_214"
                     /note="Organ: Lung; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with Not I and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is GATAAGGCCA. Tissue was provided by Mary Hendrix."
ORIGIN

  Query Match           15.8%;  Score 315.8;  DB 6;  Length 703;
  Best Local Similarity   73.4%;  Pred. No. 5.2e-40;
  Matches  509;  Conservative    0;  Mismatches  152;  Indels   32;  Gaps    7;

Qy       1076 TCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGT 1135
              | ||||   || ||||| | |||||||||| || |||  ||| | | |||||| ||||||
Db        703 TATTTTTATTTTCCAGTACCTATAAAAGTTATGTTTAATCTATATTATAGTCTATTAAGT 644

Qy       1136 GTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAA 1195
              ||||||||  |||| ||| | ||   |||||| ||||||| ||||| |||||| ||||
Db        643 TTGCAATAGCATTATATCTGACAATGTACATACTTTAATTTAAAAATATTTTATTGCTAA 584

Qy       1196 AAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTC 1255
              || |||||||| || || | ||  |||||||||       || |||||||  ||  |||
Db        583 AAGATGCTAACGATAATCTAAGTCTTCAGTGAGT-----CTTTGTTGCTGGAAGACAGTC 529

Qy       1256 TTTTCTTATTGATGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGG 1304
              ||  || | || |||  | |           | | |   | |||||||| |||| ||||
Db        528 TTGCCTCACTGTTGATGGCTGCTGAATGATCGGTGTGGTGGTTGCTGAAGGTTAGTGTGG 469

Qy       1305 CTGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCT 1359
              || | |   |||||| ||     |||||| |||| |||| |||| ||||||||||
Db        468 CTCTAGAGATTTCTTTAAATGAGACAACAAGGAAGGCTGCAACATCAACTGACTCTTCCT 409

Qy       1360 TTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGA 1419
              |||  ||||||||| ||| |||  |||| |||| || |||||| |||||  |||||| ||
Db        408 TTC-CGAAAGATTTATCTGTAGCATGTGTTGCTGTTGGATAGCCTTTTACCCACAGTGGA 350

Qy       1420 ACTTCTTTGAAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTT 1478
              || |  || ||||||| | |||||||||||||||| |||||     ||| |  | |||||||
Db        349 ACATTCTTCAAAATTGAAGTCAATCCTCTCAAACTCTGCTGCTGTTTTATCAACTAAGTT 290

Qy       1479 AATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGG 1538
              || |||||||||  | ||| |||||||||||||||||  ||||||  ||||||||||||
Db        289 TATGTAATATTCTAAGTCCTTTGTTGTCATTTCAACAATGGTCACAGGATCTTCACCAGG 230

Qy       1539 AGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCT 1598
              |||||  |||||||||||     |  |    |  |||| |  ||  ||||| || ||||
Db        229 AGTAGTTTCCATCTCAAGAAACCACTT----TTGTGGCTCTTATATTAGAAGCAACTCCT 174

Qy       1599 CATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTT 1658
              ||| |||||| |||||||||||||||||||||||| ||||||  | | ||||      |
Db        173 CATTTGTTCACGTTTTATCATGAGATTGCAGCAATTCAGTCACATATGCAGG-----TTC 119

Qy       1659 CACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGA 1718
              ||||| ||| || |||| ||||||| |||| |||||||||  ||| |||| |||| |||
Db        118 CACTTCTAAGTCTAGTTATCTTGCTATTTCCACCACATCTGCAGTTACTTCGTCCACTGA 59

Qy       1719 AGTCTTGAACCTCTCCAAGTCATCCATGAGGGT 1751
              |||||||||||| ||| |  |||||||| |||||
Db         58 AGTCTTGAACCCCTCAAGATCATCCATCAGGGT 26


RESULT 30
CG734253/c
LOCUS       CG734253                822 bp    DNA     linear   GSS 20-OCT-2003
DEFINITION  RP11-57I17, T7 RPCI-11 Human Male BAC Library Homo sapiens genomic
            clone RP11-57I17, genomic survey sequence.
ACCESSION   CG734253
VERSION     CG734253.1  GI:37776745
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 822)
  AUTHORS   Zhao,S. and Yonescu,R.
  TITLE     Direct Submission of BAC End Sequence of CCAP clones
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Yonescu, R
            NCI/NIH
            9000 Rockville Pike, Bethesda, MD 20892, USA
            Tel: 301-402-2008
            Fax: 301-402-1204
            Email: yonescur@pop.nci.nih.gov
            Plate: 57  row: I  column: 17
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .822
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

/clone="RP11-57I17"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

ORIGIN

Query Match 15.7%; Score 314.8; DB 10; Length 822;
Best Local Similarity 69.3%; Pred. No. 7.3e-40;
Matches 592; Conservative 0; Mismatches 212; Indels 50; Gaps 10;

```
Qy        979 CAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATG 1038
              ||| ||| |||    | |||     |||||  |||||| |||||||||||||| ||| ||||
Db        822 CAGACATTCCTCAAATATATGCAGGGTTCAGTTCCAGACCACCACAATAAAACACATATC 763

Qy       1039 CAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATAT 1098
                |  |  ||              |  |||       | ||       || ||| |  | |
Db        762 ACAACACATGA-------------GTCTCATGAGTATTTTGATATCTCTCAGAGAGTCT 717

Qy       1099 AAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAA 1158
              |||||||| || |||||||| |  ||||||| ||||||||| |||||  ||||||||||
Db        716 AAAAGTTGTGTTTATACTATATGGTAGTCTATTAAGTGTGAAATAGCATTATGTCTAAAC 657

Qy       1159 AAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGC 1218
              |||  ||||| ||||    |||| | ||||||  |||||||| ||||  || || ||| |
Db        656 AAAGTCATACTTTAA----AAAAAGGTTTATTGATAAAAAATACTAAAGATTATCTGA-C 602

Qy       1219 ATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT--- 1275
               |||| | | ||||||||||||||| ||||||| ||||||  || |  | |||  | |
Db        601 CTTCAATAATTTGTAATCTTTTTGTTGGTGGAGGGTCTTGCCTCAAGGTTGATGGCTGCT 542

Qy       1276 --------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAAC 1327
                     ||| |   | ||||||||    |||||||||||||| ||||||||      |
Db        541 GACTGATCAGGGTGGTGGTTGCTGAAGGCCGGGGTGGCTGTGGCAATTTCTTAAGCTGAG 482

Qy       1328 AGTG----AAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTG 1383
              |  |    || |||||  |||   | ||||||||||||||  |||||||  | |||||
Db        481 ACAGCAATAAAATTGCTGTATTGATGGACTCTTCCTTTCACAAAAGATT--TTTCTAGCA 424

Qy       1384 TGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATC 1443
              ||  ||||| ||||||| |||||||  | |    ||||| |||| |||||||||    |
Db        423 TGCAATGCTGTTTGATAACATTTTACCCTC---TGAACTGCTTTCAAAATTGGAGTTAAT 367

Qy       1444 CTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTT 1503
              |    |  |||||      || |  |||||||||  | ||||||||| |||| ||||
Db        366 CCTCCCGATACTGCTGCTGCTTTATCACCTAAGTTTTTGTAATATTCTGTATCCTTTGTG 307

Qy       1504 GTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAG 1563
              |||||||  |||||  |||      || |||||||||  | | |||| ||||       |
Db        306 GTCATTTTCACAATGCTCATGTCATCATCACCAGGAAGACACTCCACCTCA---AAAAAA 250

Qy       1564 ATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGA 1623
                  ||| |||| || ||| ||||| || |||||||||||||||||| ||||||
Db        249 CCACTTTCTCTGCTGATTCATGAGAAGCAACTCCTCATCTGTTCAAGTTTGATCATGAAT 190

Qy       1624 TTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCT 1683
              |||  ||||| |||||| |||||||||     || ||||| |||||  ||||||||||
Db        189 TTGTCGCAATTCAGTCACGTCTTCAGG-----CTCCACTTCTAATTGTAGTTCTCTTGCG 135

Qy       1684 GTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCC 1743
                |||  |||||||||    |||  |||||||  |||||||||||||||||| |||||| ||
Db        134 ATTTTCACCACATCTTCAGTTATTTCCTCCGCTGAAGTCTTGAACCTCTCAAAGTCAGCC 75

Qy       1744 ATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCC 1800
              |||||||||||||  |||||| ||||| |||| ||||| || ||||||| |   | |||
Db         74 ATGAGGGTTGGAATTAACTTCTACCAAACTCCTATTAATGTTGATATTTTAAGCTCTTCC 15

Qy       1801 CATGAATCATGAAT 1814
              ||||||||| ||||
Db         14 CATGAATCACGAAT 1

.
```

RESULT 31
BU740723

```
LOCUS       BU740723                668 bp    mRNA    linear   EST 10-OCT-2002
DEFINITION  UI-E-EJ0-ail-h-12-0-UI.s1 UI-E-EJ0 Homo sapiens cDNA clone
            UI-E-EJ0-ail-h-12-0-UI 3', mRNA sequence.
ACCESSION   BU740723
VERSION     BU740723.1  GI:23681014
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 16-652, >TIGGER1#DNA/MER2_type (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EJ0-ail-h-12-0-UI"
                     /tissue_type="fetal eyes, lens, eye anterior segment,
                     optic nerve, retina, Retina Foveal and Macular, RPE and
                     Choroid"
                     /dev_stage="fetal and adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-EJ0"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-EJ0 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tags for this library are: fetal eyes,
                     AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment,
                     AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina
                     Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
                     library was created for the program, Gene Discovery in the
                     Visual System, supported by National Eye Institute (NEI).
                     TAG_TISSUE=human retina
                     TAG_LIB=UI-E-EJ0
                     TAG_SEQ=CCGCG"

ORIGIN

  Query Match 15.6%; Score 312.6; DB 5; Length 668;
```

```
  Best Local Similarity   76.4%;  Pred. No. 1.7e-39;
  Matches  485;  Conservative    0;  Mismatches  119;  Indels   31;  Gaps    7;

Qy       1354 CTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCAC 1413
             ||| ||||| |||||||||||||| |||  || || ||| ||||| || ||||||   |
Db         33 CTTTATTTCAGGAAAGATTTCTCTGTAGCATGCGACGCTGTTTGACAGAATTTTACCAAT 92

Qy       1414 AGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACC 1472
             ||||| |||||||| ||||||||| | || ||||||||||||||      || || | |
Db         93 AGTAGGACTTCTTTCAAAATTGGAGTTAAACCTCTCAAACCCTGCCACTACTTTACCAAC 152

Qy       1473 TAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTC 1532
             |||||| ||   |||||||||||||  |||||| |||||||||| |||||||  ||||
Db        153 TAAGTTTATGATATATTCTGAATCCTGTGTTGTAATTTCAACAATATTCACAGCATCTTT 212

Qy       1533 ACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAA 1592
             |  ||||||||||||||| |||       |  |    ||||  |||||||||| ||||||
Db        213 AGTAGGAGTAGATTCCATTTCAAGAAACCACCT----TCTTCACTCATCCATAGGAAGAA 268

Qy       1593 ATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCC 1652
             || |||||||  || |||||||| |||||||| |||||| ||||||    |||||||
Db        269 ATACCTCATCCATTAAAGTTTTACCATGAGATTGTAGCAATTCAGTCACAACTTCAGG-- 326

Qy       1653 TCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTC 1712
                | |||||  ||||  | |||||||||| |||| ||||||||    |||  ||||||
Db        327 ---TTCCACTTCGAATTATATTTCTCTTGCTATTTCCACCACATCCACAGTTATTTCCTC 383

Qy       1713 CATTGAAGTCTTGAACCTCTC----------CAAGTCATCCATGAGGGTTGGAATCGA 1760
             || ||| |||||||||| |||           |||||||||| | ||||||||||  |
Db        384 CACTGATGTCTTGAACCCCTCCATGAAGGGGTTCAAGTCATCCATAA-GGTTGGAATCAA 442

Qy       1761 CTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATCATGAATGTT 1817
             ||||| ||||  |||||||||| ||  |||||||   |||||||||||||  ||||||
Db        443 CTTCTCCCAAGCTCCTGTTAATGTTGGTATTTTGACTGCCTCCCATGAATCACAAATGTT 502

Qy       1818 CTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGA 1877
             |||||  ||| || |||| || |||||||||| ||||||||||||||||||| | |||||
Db        503 CTTAACAGCATCTAGAATAGTAAATCCTTTCCAGAAGGTTTTCAATTTACTTTGCCCAGA 562

Qy       1878 TCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTT 1937
             |||||      |||||| |||| || ||| ||| |||||||||   ||||||| || |
Db        563 TCCAT-----GAGAGGAATCACTATCTATGACAGCTATAGCCTTACAAAATGTATGTCAT 617

Qy       1938 CAATAATAAGGCTTGAAAGTTGAAATTACTCCTTG 1972
               || |||||| |||||||||  |||| ||||||||
Db        618 AAACAATAAGACTTGAAAGTCAAAATGACTCCTTG 652


RESULT 32
BX094980/c
LOCUS       BX094980                774 bp    mRNA    linear   EST 04-FEB-2003
DEFINITION  BX094980 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGp998G1990 ; IMAGE:112554, mRNA sequence.
ACCESSION   BX094980
VERSION     BX094980.1  GI:27827410
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
            Radelof,U., Schneider,D. and Korn,B.
  TITLE     Human UnigeneSet - RZPD3
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD; IMAGp998G1990.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
            Human UnigeneSet - RZPD3 (RZPDLIB No.972)
            http://www.rzpd.de/CloneCards/cgi-
            bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998G1990 ; IMAGE:112554"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

  Query Match           15.6%;  Score 311.8;  DB 5;  Length 774;
  Best Local Similarity   72.3%;  Pred. No. 2.2e-39;
  Matches  568;  Conservative    0;  Mismatches  178;  Indels   40;  Gaps   11;

Qy       1113 TACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAA---CACATACC 1169
             ||| | |||| ||||| |||| |||||| | |  |||||| |||||  |    ||||| |
Db        774 TACCATACTGGAGTCTTTTAAATGTGCACTGGCATTATGTTTAAAANCAATGTACATATC 715

Qy       1170 TTAATTTTAAAATG-CTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAG 1228
             |||||||||||| | |||||  | |||||  |||||   ||||| ||||| |||||  ||
Db        714 TTAATTTTAAAAAGACTTTACAAGTAAAACTTGCTAGAGATCATCTGAGCCTTCAGCAAG 655

Qy       1229 TTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CG 1277
             |  |||||||||||| | |||| |||||  ||   || ||  | |             |
Db        654 TCATAATCTTTTTGCAGATGGAGGGTCTTGCCTCGATGTTGTTGGTTGTTGACTGATCAG 595

Qy       1278 GGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATT 1337
             || |   | |||||||| |  |||||||||||||| ||||||||| |  ||  ||
Db        594 GGTGGTGGTTGCTGAAGGTAGGGGTGGCTGTGGCAATTTCTTAAAATA--AGACAACGAA 537

Qy       1338 GCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTG 1397
             |        |||||||| ||||||||| |||||||||| |||   |  || || ||||
Db        536 GTTTGCCACATTGACTCTCCCTTTCATGGAAGATTTCTCTGTAGGAGGCAATACTGTTTG 477

Qy       1398 ATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACCCTGC 1457
             |||||| |||| |||||||||||||| ||  ||| | |  ||| |||||| |||||||||
Db        476 ATAGCA-TTTACCCACAGTAGAACTT-TTCAAAATTGGAGTCAGTCCTCTAAAACCCTGC 419

Qy       1458 TCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAAT 1517
             |      || |  | |||||||  ||  ||||| ||||| |||| |||||||| ||||
Db        418 TGCTGCTTTATCAACTAAGTTTATGAAACATTCTAAATCCTTTGTCATCATTTCAGCAAT 359

Qy       1518 TTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCT 1577
              |||||||  ||||||    ||||||||| ||  |||        |       | ||||||
Db        358 GTTCACAGCTTCTTCA----GAGTAGATTTCACTTCAA----AAAACATCTTTATTTGCT 307

Qy       1578 CATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAG 1637
             | ||||||||| |||||||||  || |  |||||||||||||||||||||||||| |||
Db        306 TAGCCATAAGAAGCAATTCCTCATCCATTTAGATTTTATCATGAGATTGCAGCAATTCAG 247

Qy       1638 TCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATC 1697
             |||  | |||||     | ||||||| |||||| |||||||||||  ||| || ||||||
Db        246 TCACATTTTCAG-----ATTTCACTTCTAATTCAAGTTCTCTTGCCATTTATATCACATC 192
```

```
Qy       1698 TGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAAT 1757
Db        191 TGCAG----TTTCACCACTGAAGTCTTAAACCCCTCAAAATCATTCTTGAGGGCTGGAAT 136

Qy       1758 CGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTT 1817
Db        135 CAACGTTTGCCAAACTCCTCTTAATGCTGATATTTTGACCTCC----TCCCAGGAATGTT 80

Qy       1818 CTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGA 1877
Db          79 CTTCATGGCATCTAAAATGGTAAATCCTTTCCAGAAGGTTTTCAATTTACTTTTCCGAGA 20

Qy       1878 TCCATC 1883
Db          19 TTCATC 14
```

RESULT 33
CR736370/c
LOCUS CR736370 774 bp mRNA linear EST 10-DEC-2004
DEFINITION CR736370 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGp998B2495 ; IMAGE:114359 5', mRNA sequence.
ACCESSION CR736370
VERSION CR736370.1 GI:51586310
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 774)
AUTHORS Ebert,L., Heil,O., Hennig,S., Korn,B., Neubert,P., Partsch,E., Peters,M., Radelof,U. and Schneider,D.
TITLE I.M.A.G.E. cDNA Clone Collection
JOURNAL Unpublished (2004)
COMMENT Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Email: www.rzpd.de
RZPD; IMAGp998B2495.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 100
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD; contact RZPD (clone@rzpd.de) for further information. Seq primer: M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES Location/Qualifiers
source 1. .774
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGp998B2495 ; IMAGE:114359"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen 1NFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

Query Match 15.6%; Score 311.8; DB 7; Length 774;
Best Local Similarity 72.3%; Pred. No. 2.2e-39;
Matches 568; Conservative 0; Mismatches 178; Indels 40; Gaps 11;

```
Qy       1113 TACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAA---CACATACC 1169
Db        774 TACCATACTGGAGTCTTTTAAATGTGCACTGGCATTATGTTTAAAANCAATGTACATATC 715

Qy       1170 TTAATTTTAAAATG-CTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAG 1228
Db        714 TTAATTTTAAAAAGACTTTACAAGTAAAACTTGCTAGAGATCATCTGAGCCTTCAGCAAG 655

Qy       1229 TTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CG 1277
Db        654 TCATAATCTTTTTGCAGATGGAGGGTCTTGCCTCGATGTTGTTGGTTGTTGACTGATCAG 595

Qy       1278 GGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATT 1337
Db        594 GGTGGTGGTTGCTGAAGGTAGGGGTGGCTGTGGCAATTTCTTAAAATA--AGACAACGAA 537

Qy       1338 GCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTG 1397
Db        536 GTTTGCCACATTGACTCTCCCTTTCATGGAAGATTTCTCTGTAGGAGGCAATACTGTTTG 477

Qy       1398 ATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACCCTGC 1457
Db        476 ATAGCA-TTTACCCACAGTAGAACTT-TTCAAAATTGGAGTCAGTCCTCTAAAACCCTGC 419

Qy       1458 TCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAAT 1517
Db        418 TGCTGCTTTATCAACTAAGTTTATGAAACATTCTAAATCCTTTGTCATCATTTCAGCAAT 359

Qy       1518 TTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCT 1577
Db        358 GTTCACAGCTTCTTCA----GAGTAGATTTCACTTCAA----AAAACATCTTTATTTGCT 307

Qy       1578 CATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAG 1637
Db        306 TAGCCATAAGAAGCAATTCCTCATCCATTTAGATTTTATCATGAGATTGCAGCAATTCAG 247

Qy       1638 TCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATC 1697
Db        246 TCACATTTTCAG-----ATTTCACTTCTAATTCAAGTTCTCTTGCCATTTATATCACATC 192

Qy       1698 TGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAAT 1757
Db        191 TGCAG----TTTCACCACTGAAGTCTTAAACCCCTCAAAATCATTCTTGAGGGCTGGAAT 136

Qy       1758 CGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTT 1817
Db        135 CAACGTTTGCCAAACTCCTCTTAATGCTGATATTTTGACCTCC----TCCCAGGAATGTT 80

Qy       1818 CTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGA 1877
Db          79 CTTCATGGCATCTAAAATGGTAAATCCTTTCCAGAAGGTTTTCAATTTACTTTTCCGAGA 20

Qy       1878 TCCATC 1883
Db          19 TTCATC 14
```

RESULT 34
BM715720
LOCUS BM715720 724 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-EJ0-ahj-k-13-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone UI-E-EJ0-ahj-k-13-0-UI 5', mRNA sequence.
ACCESSION BM715720
VERSION BM715720.1 GI:19028978
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 724)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 8-724, >TIGGER1#DNA/MER2_type (matched compliment)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EJ0-ahj-k-13-0-UI"
                     /tissue_type="fetal eyes, lens, eye anterior segment,
                     optic nerve, retina, Retina Foveal and Macular, RPE and
                     Choroid"
                     /dev_stage="fetal and adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-EJ0"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-EJ0 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tags for this library are: fetal eyes,
                     AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment,
                     AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina
                     Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
                     library was created for the program, Gene Discovery in the
                     Visual System, supported by National Eye Institute (NEI)."
ORIGIN

  Query Match             15.6%;  Score 311.6;  DB 3;  Length 724;
  Best Local Similarity   72.8%;  Pred. No. 2.4e-39;
  Matches  537;  Conservative    0;  Mismatches  164;  Indels   37;  Gaps    9;

Qy       1106 TTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC--- 1162
Db          3 TTTTATAGGCTGTAATGTAATCTACTAAGTATGCAGTAGCATTATGTCTAAAAAGCCAAT 62

Qy       1163 --ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCAT 1220
Db         63 GTATGTACCTTAATTTTAAAATACTTTGTTGCTAAAAAATGCTAATGATCATCTGAGTGT 122

Qy       1221 TCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTG------- 1273
Db        123 TCAGTGAATTGCCATCTTTTTGCTGGGGGAGGGTCTTGCCTCAATGTTGGTGGCTGCTGA 182

Qy       1274 ----ATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACA--AC 1327
Db        183 CTAATTAGGGTGGTGGCTGCTGGAGTTTGGGGTTGCTGTGGCAATTTCTTAAAATAAGAC 242

Qy       1328 AGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTG 1387
Db        243 AATGAAGTTTGTCACATCAATTGACTTCCATT----CAAAAGACTTCTCTGTAGCATGTG 298

Qy       1388 ATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCT 1447
Db        299 TTGCTGTTTGATAGCAGTTTACCCACATAACATCT--TTAAAAATTAGAGTCACTCTTCT 356

Qy       1448 CAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCA 1507
Db        357 CAAACCCTGCCACTGCTTTATCAACTAAGTTGATGTCA-ATTCTAAATCCTTTGTTGTCA 415

Qy       1508 TTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGG 1567
Db        416 TTTCAACCGTGTGCACAGCATCTTCACCAGAAGTAGATTCCATCTCA----AGAAAACAC 471

Qy       1568 AATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGC 1627
Db        472 TTCCTTTGTTCATCCATAAGAAGCAACTTGTCATCCATTCATGTTTTCTCATGAGACTGC 531

Qy       1628 AGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTT 1687
Db        532 AGCAATTCAGTCCCATCTCCAGG-----CTCCACTTCCAATTCTCGTTCTCTTGCTGTTT 586

Qy       1688 CTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGA 1747
Db        587 CCACCACATCTGCAGTGAGTTCCTCCCCTGAAGTCCGGGATCGCTCAGAGTCATCCATGA 646

Qy       1748 GGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATG 1804
Db        647 GGGCTGGAATCAACATCTTCCACACTCCTGTTCATGTTGCTGTTTTGACCTCCTCCCATG 706

Qy       1805 AATCATGAATGTTCTTAA 1822
Db        707 AATCATGAATGTTCTTAA 724


RESULT 35
BX503658
LOCUS       BX503658               673 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686B01107_s1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686B01107 3', mRNA sequence.
ACCESSION   BX503658
VERSION     BX503658.1  GI:32026288
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
            Wellenreuther,R., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 3' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No r1 sequence available.
            This clone (DKFZp686B01107) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Homo sapiens"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686B01107"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             15.6%;  Score 311.4;  DB 5;  Length 673;
  Best Local Similarity   73.4%;  Pred. No. 2.6e-39;
  Matches  504;  Conservative    0;  Mismatches  151;  Indels   32;  Gaps    7;

Qy       1064 GAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGT 1123
             |||||||||||| | ||||    || |||||| | ||||| |||| || |||  ||| | | |
Db             1 GAGTCACACAATTATTTTTATTTTCCAGTACCTATAACAGTTATGTTTAATCTATATTAT 60

Qy       1124 AGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATG 1183
             ||||| |||||| ||||||||  |||| ||| | ||    |||||| |||||||| |||||
Db            61 AGTCTATTAAGTTTGCAATAGCATTATATCTGACAATGTACATACTTTAATTTAAAAATA 120

Qy       1184 CTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGC 1243
              |||||| |||||| |||||||| || || | ||  ||||||||||        || ||||
Db            121 TTTTATTGCTAAAAGATGCTAACGATAATCTAAGTCTTCAGTGAGT-----CTTTGTTGC 175

Qy       1244 TGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CGGGGGTCAGGTGCTGA 1292
              |||  ||  |||||  || | || |||  | |           | | |   | ||||||
Db            176 TGGAAGACAGTCTTGCCTCACTGTTGATGGCTGCTGAATGATCGGTGTGGTGGTTGCTGA 235

Qy       1293 AGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGCAATATCAG 1347
              || |||| |||||| | |   |||||| ||     ||||||  ||||  ||||| ||||
Db            236 AGGTTAGTGTGGCTCTAGAGATTTCTTTAAATGAGACAACAAGGAAGGCTGCAACATCAA 295

Qy       1348 TTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTT 1407
              |||||||||||||   ||| ||||| ||| |||  |||| |||| || |||||| ||||
Db            296 CTGACTCTTCCTTTC-CGAAGGATTTATCTGTAGCATGTGTTGCTGTTGGATAGCCTTTT 354

Qy       1408 ATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTA 1466
              |  |||||| |||| |  || |||||||| | |||||||||||||||| |||||    |||
Db            355 ACCCACAGTGGAACATTCTTCAAAATTGAAGTCAATCCTCTCAAACTCTGCTGCTGTTTT 414

Qy       1467 ACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGT 1526
              |  | ||||||| || ||||||||| | ||| |||||||||||||||||  ||||||
Db            415 ATCAACTAAGTTTATGTAATATTCTAAGTCCTTTGTTGTCATTTCAACAATGGTCACAGG 474

Qy       1527 GTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAA 1586
              |||||||||||||||| ||||||||||        |  |   |  | |||| |  || |
Db            475 ATCTTCACCAGGAGTAGTTTCCATCTCAAGAAACCACTT----TTGTGGCTCTTATATTA 530

Qy       1587 GAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTT 1646
              |||| || ||||||| |||||| |||||||||||||||||||||||||| ||||||  | |
Db            531 GAAGCAACTCCTCATTTGTTCACGTTTTATCATGAGATTGCAGCAATTCAGTCACATATG 590

Qy       1647 CAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCC 1706
              ||||      | ||||| ||| || |||| ||||||| |||| |||||||||  ||| |
Db            591 CAGG-----TTCCACTTCTAAGTCTAGTTATCTTGCTATTTCCACCACATCTGCAGTTAC 645

Qy       1707 TTCCTCCATTGAAGTCTTGAACCTCTC 1733
              ||| |||| |||||||||||||| |||
Db            646 TTCGTCCACTGAAGTCTTGAACCCCTC 672


RESULT 36
CE455208
LOCUS       CE455208                671 bp    DNA     linear   GSS 27-SEP-2003
DEFINITION  tigr-gss-dog-17000319501415 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE455208
VERSION     CE455208.1  GI:36749305
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .671
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             15.6%;  Score 311.2;  DB 10;  Length 671;
  Best Local Similarity   72.6%;  Pred. No. 2.8e-39;
  Matches  510;  Conservative    0;  Mismatches  158;  Indels   34;  Gaps    7;

Qy       1287 TGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAATATCA 1346
              |||   || || || ||| |||||||| ||||||||   |||    | | |||    ||
Db             1 TGCGACAGGTTGGGTTGGGTGTGGCAATTTCTTAATGTAAC----ATGTTTGGTGCATTG 56

Qy       1347 GTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTT 1406
              |||| |||||||||||||    ||||||| | |  || | |||| ||||| |||||||
Db            57 ATTGAGTCTTCCTTTCATGATCATTTTCTCTGTGGCATGCGGTGCTGTTTGAAAGCATTT 116

Qy       1407 TATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACCCTGCTCTGCTTTA 1466
              ||  |||||| | || ||||   || | | ||||||||||||||  |||        ||
Db           117 TACACACAGTGGCACGTCTTCAGAATTGGAATCAATCCTCTCAAAATCTGTAGACGCTTT 176

Qy       1467 ACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGT 1526
              |  | | ||    || || |||| | ||| | |||||| |||||||||||| |||||||
Db           177 ATCAACCAACCATATGTAGTATTTTAAATTCTTTGTTGCCATTTCAACAATCTTCACAGC 236

Qy       1527 GTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAA 1586
              |||||||||||| | |||  ||| ||       || |      |  |||||||  ||||
Db           237 ACCTTCACCAGGAGGAAATTTTATCCCA-----AGAAACCATGTTATTGCTCATTTATAA 291

Qy       1587 GAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTT 1646
              |||| || || |||||| || ||| |||||||||| ||          ||||  ||||
Db           292 GAAGCAACTCTTCATCTCTTGAAGCTTTATCATGAGTTT----------GTCACCTCTT 340

Qy       1647 CAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCC 1706
              |||||  |||||||     ||||| ||||||||||||| |||||  |||||||  ||| |
Db           341 CAGGCTCCACTTCA-----AATTCTAGTTCTCTTGCTGTCTCTACTGCATCTGTAGTTAC 395

Qy       1707 TTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTT 1766
              |||||||| || |||||||||||||| |||||| |||||| |||||||||| |||| ||
Db           396 TTCCTCCACTG-AGTCTTGAACCTCTCAAAGTCAGCCATGATGGTTGGAATCAACTTTTT 454

Qy       1767 CCAAATTCCTGTTAATATTTATATTTTGACC---TCCCATGAATCATGAATGTTCTTAAT 1823
              ||||| ||||||| || || ||||||||||    |  ||||||||    | ||||||||
Db           455 CCAAACTCCTGTTTATGTTGATATTTTGACCACTTTTCATGAATCGCAAGTGTTCTTACC 514

Qy       1824 GGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATC 1883
              | || || |||| ||| |  ||||||| |||||||||| ||| ||| || |||||||||
Db           515 GACATCTAGAATAGTGGACTCTTTCCAGAAGGTTTTCACTTTGCTTTGTTCAGATCCAT- 573
```

```
Qy       1884 CATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAA 1943
              |||| ||  |||| || |||  || | |||||||||   ||||||||||| |||||
Db        574 ----CAGAAGAATCACTGTCTATGGAAGCTGTAGCCTTACTAAATGTATTTCTTAAATAA 629

Qy       1944 TAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGC 1985
              |||    ||||  ||||||||||||||||||||||||   ||||
Db        630 TAATATGTGAATATTGAAATTACTCCTTGATCCATGGGCTGC 671


RESULT 37
BM978279
LOCUS       BM978279                690 bp    mRNA    linear   EST 21-FEB-2003
DEFINITION  UI-CF-EC1-aeb-n-17-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
            UI-CF-EC1-aeb-n-17-0-UI 3', mRNA sequence.
ACCESSION   BM978279
VERSION     BM978279.1  GI:19597547
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com) or from Open Biosystems
            (www.openbiosystems.com).
             The following repetitive elements were found in this cDNA
            sequence: 20-690, >TIGGER1#DNA/MER2_type (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-EC1-aeb-n-17-0-UI"
                     /tissue_type="Lung"
                     /dev_stage="Adult and Fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-CF-EC1"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-EC1 is a normalized cDNA library containing the
                     following tissue(s): Normal lung from adult and from fetal
                     day 64, day 87, week 19 and week 42. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site.  Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AAGTGCTTAC.
                     TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
                     and 380-383
                     TAG_LIB=UI-CF-EC1
                     TAG_SEQ=AAGTGCTTAC"
ORIGIN

  Query Match           15.5%;  Score 310.6;  DB 3;  Length 690;
  Best Local Similarity   73.5%;  Pred. No. 3.5e-39;
  Matches  502;  Conservative    0;  Mismatches  149;  Indels   32;  Gaps    7;

Qy       1064 GAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGT 1123
              |||||||||||  | ||||   || ||||| | ||||||||||  || |||  ||| | | |
Db         20 GAGTCACACAAATATTTTTATTTTCCAGTACCTATAAAAGTTATGTTTAATCTATATTAT 79

Qy       1124 AGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATG 1183
              ||||| |||||| ||||||||  |||| ||| | ||   |||||| ||||||| |||||
Db         80 AGTCTATTAAGTTTGCAATAGCATTATATCTGACAATGTACATACTTTAATTTAAAAATA 139

Qy       1184 CTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGC 1243
              |||||| |||||| |||||||| || || | ||  |||||||||       || ||||
Db        140 TTTTATTGCTAAAAGATGCTAACGATAATCTAAGTCTTCAGTGAGT-----CTTTGTTGC 194

Qy       1244 TGGTGGAAGGTCTTTTCTTATTGATGACTGAT----------CGGGGGTCAGGTGCTGA 1292
              |||  ||  |||||  || | || |||  | |           | | |   | ||||||
Db        195 TGGAAGACAGTCTTGCCTCACTGTTGATGGCTGCTGAATGATCGGTGTGGTGGTTGCTGA 254

Qy       1293 AGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGCAATATCAG 1347
              || |||| |||||| | |   |||||| ||     ||||||  ||||  ||||| ||||
Db        255 AGGTTAGTGTGGCTCTAGAGATTTCTTTAAATGAGACAACAAGGAAGGCTGCAACATCAA 314

Qy       1348 TTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTT 1407
              |||||||||||||  |||||||| ||| |||  |||| |||| || |||||  ||||
Db        315 CTGACTCTTCCTTTC-CGAAAGATTTATCTGTAGCATGTGTTGCTGTTGGATAGCCTTTT 373

Qy       1408 ATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTA 1466
              |  |||||| |||| |  || ||||||| | ||||||||||||||| |||||     |||
Db        374 ACCCACAGTGGAACATTCTTCAAAATTGAAGTCAATCCTCTCAAACTCTGCTGCTGTTTT 433

Qy       1467 ACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGT 1526
              |  | ||||||| || |||||||| | ||| |||||||||||||||||  ||||||
Db        434 ATCAACTAAGTTTATGTAATATTCTAAGTCCTTTGTTGTCATTTCAACAATGGTCACAGG 493

Qy       1527 GTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAA 1586
              |||||||||||||||| |||||||||       |  |    |  | |||| |  || |
Db        494 ATCTTCACCAGGAGTAGTTTCCATCTCAAGAAACCACTT----TTGTGGCTCTTATATTA 549

Qy       1587 GAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTT 1646
              |||| || ||||||| |||||| |||||||||||||||||||||||||| ||||||  | |
Db        550 GAAGCAACTCCTCATTTGTTCACGTTTTATCATGAGATTGCAGCAATTCAGTCACATATG 609

Qy       1647 CAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCC 1706
              ||||       | ||||| ||| || |||| ||||||| |||| ||||||||||  ||| |
Db        610 CAGG-----TTCCACTTCTAAGTCTAGTTATCTTGCTATTTCCACCACATCTGCAGTTAC 664

Qy       1707 TTCCTCCATTGAAGTCTTGAACC 1729
              ||| |||| |||||||||| ||
Db        665 TTCGTCCACTGAAGTCTTGACCC 687


RESULT 38
BP280947/c
LOCUS       BP280947                537 bp    mRNA    linear   EST 16-SEP-2004
DEFINITION  BP280947 Sugano cDNA library, KG-1-C Homo sapiens cDNA clone
            KGR09271, mRNA sequence.
ACCESSION   BP280947
VERSION     BP280947.1  GI:52194679
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
```

```
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
   PUBMED   15342556
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="KGR09271"
                     /cell_line="KG-1-C"
                     /clone_lib="Sugano cDNA library, KG-1-C"
                     /note="glioma"
ORIGIN

  Query Match             15.5%;  Score 310.4;  DB 3;  Length 537;
  Best Local Similarity   80.2%;  Pred. No. 3.9e-39;
  Matches  417;  Conservative    0;  Mismatches   91;  Indels   12;  Gaps    4;

Qy       1319 TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTC 1378
              ||| ||| | |||||| ||      || ||||||||||||||||||||||||||||||||
Db        511 TAAGACATCGGTGAAGCTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTG 452

Qy       1379 TAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA- 1437
              |||  || |||||| ||||||||||||||    |||||||| ||||||| |||||||||
Db        451 TAGCATGCGATGCTGTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAG 392

Qy       1438 TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCC 1497
              ||||||||||||||||||||     || |  | ||||||| || |||||||||| |||||
Db        391 TCAATCCTCTCAAACCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCC 332

Qy       1498 ATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTT 1557
              ||||||||||||||||||| |||||||  |||||||||| ||||||||||||||||||
Db        331 TTTGTTGTCATTTCAACAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGA 272

Qy       1558 CCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATC 1617
                |  |    ||||||| |||||||||||  || ||||||||||| ||||||||||
Db        271 AACCACTT----TCTTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATC 216

Qy       1618 ATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCT 1677
              ||||||||||||||| |||| |  ||||||||     ||  |||| |||||| ||||||
Db        215 ATGAGATTGCAGCAATTCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCT 161

Qy       1678 CTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAG 1737
              |||||  |||| |||||||||    || ||||||||| ||||||||||||||| |||  ||
Db        160 TTTGCTATTTCCACCACATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAG 101

Qy       1738 TCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACC 1797
              |||||  |||||||| ||||||  |||||||||||| |||||||||| || | | |||||||
Db        100 TCATCAATGAGGGTGGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACC 41

Qy       1798 TCCCATGAATCATGA--ATGTTCTTAATGGCACCTGGAAT 1835
              |||    |   || |  | ||| |||||||||| ||  |||
Db         40 TCCTCCCATGAATCACGAAGTTTTTAATGGCATCTAAAAT 1


RESULT 39
BZ771573/c
LOCUS       BZ771573               1011 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  mct78a04.g6 HFOSMID004 Homo sapiens genomic, genomic survey
            sequence.
ACCESSION   BZ771573
VERSION     BZ771573.1  GI:28945257
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1011)
  AUTHORS   Cook,L., Delehaunty,K., Fewell,G., Fulton,L., Magrini,V.,
            Mardis,E., Miner,T., Nash,W., Williams,D. and Wilson,R.K.
  TITLE     Homo sapiens Fosmid End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: mct78  row: a  column: 04
            Class: fosmid ends
            High quality sequence start: 102
            High quality sequence stop: 526.
FEATURES             Location/Qualifiers
     source          1. .1011
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="HFOSMID004"
                     /note="Vector: pcc01fos; Site_1: Eco72i; Human whole
                     genome fosmid library was prepared at Washington
                     University Genome Sequencing Center. DNA was sheared for
                     blunt-ended ligation into pcc01fos inducible vector. DNA
                     was ordered from Coriell Cell Repository's DNA
                     polymorphism discovery resource."
ORIGIN

  Query Match             15.3%;  Score 306.6;  DB 9;  Length 1011;
  Best Local Similarity   77.0%;  Pred. No. 1.4e-38;
  Matches  441;  Conservative    0;  Mismatches  119;  Indels   13;  Gaps    5;

Qy       1319 TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTC 1378
              ||| |||||| |||||    || | ||    ||||||||||||||||||||| |||||||
Db        677 TAAGACAACATTGAAGTGAGCCACATGGATTGACTCTTCCTTTCATGAAAGGTTTCTCTG 618

Qy       1379 TAGTGTGTGATGCTTTTTGATAGCATTTTA--TGCACAGTAGAACTTCTTTGAAAATTGG 1436
              ||| || |||||||  |||||||||||||||  | |||||||  ||    |||| |||||||
Db        617 TAGCGTATGATGCTGTTTGATAGCATTTTACCTACACAGTATAATGCCTTTCAAAATTGC 558

Qy       1437 A-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAAT 1495
              | ||| || |||||||  ||||||      || |  | |||| || || | |  |||| |||
Db        557 AGTCAGTCTTCTCAAATCCTGCTGCCACTTTATCAACTAACTTCATGTGAAGTTCTAAAT 498

Qy       1496 CCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCAT 1555
              || || |||   |||||||||| |||||||  | |||||||  |||||  ||||    | |
Db        497 CCTTTATTGCTGTTTCAACAATGTTCACAGCATTTTCACCAAAAGTAGTTTCCTGATTCT 438

Qy       1556 TTCCTGAGATGGA--ATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTT 1613
              |  | |  |   |  |||||||||||||| ||||||  |  | ||||||||||  || ||
Db        437 TCTCAGGAAACCACTTTCTTTGCTCATCCACAAGAAGCAGCTTCTCATCTGTTTGAGATT 378

Qy       1614 TATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAG 1673
              ||||||||||  ||| ||| ||||||  |||||||||  ||||||     |||||| ||
Db        377 GGTCATGAGATTTCAGTAATTCAGTCACATCTTCAGGCTCCACTTC-----TAATTCTAG 323

Qy       1674 TTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTC 1733
              ||||||||||  || ||||||||||   ||   | |||||| ||||||||||||||| |||
Db        322 TTCTCTTGCTACTTTTACCACATCTGCAGTGACCTCCTCCACTGAAGTCTTGAACCCCTC 263

Qy       1734 CAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTT 1793
              |||||||||   ||||||  || ||||  ||| |||||||  ||||||||||  || ||||||
Db        262 CAAGTCATGTGTGAGGGGTGCAATCAGCTTTTTCCAAACTCCTGTTAATGTTGATATTTG 203

Qy       1794 GA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
              ||   |||||||||||||  ||| |||||||||||| || |||||  | ||||||||||||
Db        202 GACCTCCTCCCATGAATCACAAATCTTCTTAATGGCATCTAGAATGATTAATCCTTTCCA 143

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATC 1883
              |||||||||||| ||||  | ||||||||||||
Db        142 GAAGGTTTTCAATATACTCTGCCCAGATCCATC 110
```

```
RESULT 40
CC765346
LOCUS       CC765346               723 bp    DNA     linear   GSS 27-JUN-2003
DEFINITION  CH240_57C13.TV CHORI-240 Bos taurus genomic clone CH240_57C13,
            genomic survey sequence.
ACCESSION   CC765346
VERSION     CC765346.1  GI:32311849
KEYWORDS    GSS.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,
            Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,
            Womack,J.E., de Jong,P.J. and Lewin,H.A.
  TITLE     Bovine BAC end sequences from CHORI-240 library
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CH240_57C13.TJ
            Contact: Harris Lewin
            Department of Animal Sciences
            University of Illinois at Urbana Champaign
            1201 W. Gregory Dr., Urbana, IL 61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
            Email: h-lewin@uiuc.edu
            Clones are derived from the bovine BAC library CHORI-240
            (http://www.chori.org/bacpac/bovine240.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/ordering_information.htm). This work
            was undertaken as part of the International Bovine BAC Mapping
            Consortium (IBBMC) by by University of Illinois at Urbana
            Champaign, USA with funds provided by grant No. AG202-34480-11828
            from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock
            Genome Sequencing Initiative)
            Plate: 57  row: C  column: 13
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="breed: Hereford"
                     /db_xref="taxon:9913"
                     /clone="CH240_57C13"
                     /sex="Male"
                     /cell_type="Blood"
                     /clone_lib="CHORI-240"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     Hereford bull Li Domino 99375; CHORI-240 Bovine BAC
                     library (Male) produced by Pieter de Jong"
ORIGIN

  Query Match             15.3%;  Score 306;  DB 9;  Length 723;
  Best Local Similarity   73.9%;  Pred. No. 1.8e-38;
  Matches  509;  Conservative    0;  Mismatches  138;  Indels   42;  Gaps    8;

Qy       1319 TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTC 1378
Db         19 TAAGACAACCANNNAGTTTGC-ATGTCGATAGATCCTCCCTTTCATGAATGATTTCTCTG 77

Qy       1379 TAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA- 1437
Db         78 TAGTGTGTGATACTGTTTGGTAGCATTTTACCTACAGTAGAACTTCTTTCAAAATTGGAG 137

Qy       1438 TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCC 1497
Db        138 TCAGTCCTCTCAAACCCCGC--------------------TGCTGTAATGTTCTAAATCT 177

Qy       1498 ATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTT 1557
Db        178 TTTATTGTCATTTCAACAATCTTCACAGTATCTTTACCAGGAGTAGTTTCTATGTCA--- 234

Qy       1558 CCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATC 1617
Db        235 -AGAAATTGCTTTCTTTTCTCCTCCATACAAAACAAGTCTTTATTCATTCTAATTTTATT 293

Qy       1618 ATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCT 1677
Db        294 AAGAGATTGCAGCAATTCAGTGACCTCTTTAGG-----CTCCAATTCTAATTCTAGTTCT 348

Qy       1678 CTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAAC---CTCTCC 1734
Db        349 TTTGCTATTTCTGCCACATCTGCAATTACTTCCTTCACTGAAGTCTTGAAGTCACTCATG 408

Qy       1735 AAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTG 1794
Db        409 AAGTCACTCATGAAGGTTGGAATCAAGCTCTTTTAAACTCCTGTTAATGTTGATATTTTG 468

Qy       1795 ACC---TCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAA 1851
Db        469 ACCTTTTTCCATGAATCATGAGTGTTCTTAATGGCATCTGAAATGAGAAATCTTTTCCAG 528

Qy       1852 AAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAG 1911
Db        529 AAGGATTTAAATTTCTTTTGCCCAGATCTAT-----CATAGGAATCACTCTCTGTTGCAG 583

Qy       1912 TTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTT 1971
Db        584 CTATAGCATTATTAAATGTATTTCTTCAGTAATAGGACTTGAGAGTCAAAATTACTCTTT 643

Qy       1972 GATCCATTTTCTGCAAAATAGATGTTGTG 2000
Db        644 GATTCATGGGCTACAGAATGGATGTTGTG 672


RESULT 41
CE027248/c
LOCUS       CE027248               746 bp    DNA     linear   GSS 24-SEP-2003
DEFINITION  tigr-gss-dog-17000344638788 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE027248
VERSION     CE027248.1  GI:35045787
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 746)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .746
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN
```

```
  Query Match             15.3%;  Score 305.6;  DB 9;  Length 746;
  Best Local Similarity   72.2%;  Pred. No. 2.1e-38;
  Matches  520;  Conservative    0;  Mismatches  164;  Indels   36;  Gaps    8;

Qy       1287 TGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACA------ACAGTGAAGATTGCA 1340
             ||||||||  ||  ||||||||||||||  |||||||| |         ||     | ||||
Db        740 TGCTGAAGGTTGGGGTGGCTGTGGCAAGTTCTTAAAATAGGACAGTGAGGTTTGTTTGCC 681

Qy       1341 ATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATA 1400
             | ||   ||||||||||||||  |||   ||||||||  |||  ||   ||||||| ||||
Db        680 ACATTGATTGACTCTTCCTTTTGTGAGCAATTTCTCTGTAGCCTGCAGTGCTTTTAGATA 621

Qy       1401 GCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACCCTGCTCT 1460
             ||||||||  ||  | |||||||  | | | |||||           |||||||||| |
Db        620 GCATTTTACACATTGGAGAACTTTTGTCATAATTG-----------CAAACCCTGCCCC 573

Qy       1461 GCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTT 1520
              || | | | | ||  || ||||||||| ||||| ||||||||||||||| | |  |
Db        572 TGCTTTATCAACCAGGTGTATGTAATATTCTAAATCCTTTGTTGTCATTTCAATAGTCCT 513

Qy       1521 CACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCAT 1580
             ||| |||| || |||| |||||||||||||||        | |    | | || ||||
Db        512 CACCGTGTTTTTACCAAGAGTAGATTCCATCTCGGGAAACCACTT----TTTCTGGTCAT 457

Qy       1581 CCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCA 1640
              ||||| | ||||| |||||  |||||||||||| |||    |||||| | |||||
Db        456 GTGTAAGAGGTAATTCTTCATCCTTTCAAGTTTTATCGTGAAGCAGCAGCAGTCCAGTCC 397

Qy       1641 TGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGT 1700
              ||||||||||  ||| ||   | |    || |    | ||||||||||||||||||
Db        396 CTTCTTCAGGCCCCACGTC----TCATCCTCATTCTCTTGGCTGTTTCTACCACATCTGC 341

Qy       1701 GGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGA 1760
              | | ||| ||||| |||||||||||||| ||| | ||||||||||||||  |||||
Db        340 AGCTACTT-CTCCACTGAAGTCTTGAACCCCTCAATGTCATCCATGAGGGTCAGAATCAC 282

Qy       1761 CTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTTCTT 1820
              |||||||||||  ||||||||||  |  |   ||   | ||||||||||||  |||||||||
Db        281 CTTCTTCCAAACTCCTGTTAATGTAGAACCTT---CTTCCCATGAATCACAAATGTTCTT 225

Qy       1821 AATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCC 1880
              |||||||  | || ||||||||||||||||| |||| ||| | |||||||   ||||||||
Db        224 AATGGCATTTAGAGTGGTGAATCCTTTCCAGAAGGGTTTTATTTTACTT-CCCCAGATCC 166

Qy       1881 ATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAA 1940
              ||      ||||||   |||| || ||| | ||||||||  ||||||||||| ||
Db        165 AT-----CAGAGGCATCACTATCTATGGCAGCTCTAGCCTTATAAAATGTATTTCTTAAA 111

Qy       1941 TAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||||  ||||||||  |||||||  ||||||||  |||    |||   ||| |||||| ||
Db        110 CAATAAAACTTGAAAGCAGAAATTATTCCTTGATTCATGGGCTGTTGAATGGATGTTTTG 51


RESULT 42
BF981537/c
LOCUS       BF981537                674 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  602309408F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4400857 5',
            mRNA sequence.
ACCESSION   BF981537
VERSION     BF981537.1  GI:12384349
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 674)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10106  row: o  column: 02
            High quality sequence stop: 664.
FEATURES             Location/Qualifiers
     source          1. .674
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4400857"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_88"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             15.3%;  Score 305.4;  DB 2;  Length 674;
  Best Local Similarity   73.0%;  Pred. No. 2.3e-38;
  Matches  465;  Conservative    0;  Mismatches  156;  Indels   16;  Gaps    5;

Qy       1368 AGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTT 1427
             | |||||||| |||  |||||||||| || ||||||||||||  |        |  | |||
Db        660 AAATTTCTCTGTAGCATGTGATGCTATTGGATAGCATTTTACCCCTGTAGAACATCTTTT 601

Qy       1428 GAAAATTGGATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATA 1487
              |||||||||  |    ||  |||||    |     |    || |||  |  |||
Db        600 CAAAATTGGAGTCAATTCCTTCAACCCCCGACGCTGCTTT--ATTAAATTTATGACGATA 543

Qy       1488 TTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTC 1547
              |||| | ||  |||||||||||||||  || |||| ||  |||||| ||| |||||||||
Db        542 TTCTAAGTCTTTTGTTGTCATTTCAATTATGTTCATAGCATCTTCATCAGAAGTAGATTC 483

Qy       1548 CATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTC 1607
              |||||||        | |  |||||||||||||||| |||||  |  ||||||||||||
Db        482 CATCTCAAGAAACCACTT----TCTTTGCTCATCCAGAAGAACCAGCTCCTCATCTGTTC 427

Qy       1608 AAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAA 1667
              |||||||||||||||||||||||||||||| ||||||  |||| ||      |||||||| |||
Db        426 AAGTTTTATCATGAGATTGCAGCAATTCAGTCACATCTTTAG------CTTCACTTCTAA 373

Qy       1668 TTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAA 1727
              ||| |||||||||||| |||| ||||||||||  ||  |||||| |||| ||||||||||
Db        372 TTCTAGTTCTCTTGCTATTTCCACCACATCTGCAGTGACTTCCTTCATTAAAGTCTTGAA 313

Qy       1728 CCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTA 1787
              | ||| |||||||||||||||||| ||||| |||||||||||| |||||||||| || |
Db        312 TCCCTCAAAGTCATCCATGAGGGTTAGAATCAACTTCTTCCAAACTCCTGTTAATCTTGA 253

Qy       1788 TATTTTGAC---CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCC 1844
              |||||||||   |||||||||||||  |||||||||||||||| || | ||||||||||
Db        252 TATTTTGACCTTCTCCCATGAATCACAAATGTTCTTAATGGCATCTAGGATGGTGAATCC 193

Qy       1845 TTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCA 1904
              |||||| |||||||||||||||||| |  || ||||||| |  |     |     |
Db        192 TTTCCAGAAGGTTTTCAATTTACTTGGCACAAATCCATCAAATAAATCATTATCTATGGT 133

Qy       1905 ATGCCAG-TTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAAT 1963
              |    ||  |  |   |       ||||||||||  || |  |     | ||||||||||
Db        132 AGCATAGCCTCAAAAATGTATTTATGTATTTCTCAAACATAAGATGTGGAAAGTTGAAAT 73

Qy       1964 TACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||||||||||||   ||||| ||| | || ||||
Db         72 TACTCCTTGATCCATGGGCTGCAGAATGGTTGCTGTG 36
```

```
RESULT 43
BU680222
LOCUS       BU680222                696 bp    mRNA    linear   EST 07-OCT-2002
DEFINITION  UI-CF-DU1-abe-n-08-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone
            UI-CF-DU1-abe-n-08-0-UI 3', mRNA sequence.
ACCESSION   BU680222
VERSION     BU680222.1  GI:23528943
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com) or from Open Biosystems
            (www.openbiosystems.com).
             The following repetitive elements were found in this cDNA
            sequence: 18-687, >TIGGER1#DNA/MER2_type (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-DU1-abe-n-08-0-UI"
                     /tissue_type="Primary Lung Epithelial Cells"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-CF-DU1"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-DU1 is a normalized cDNA library containing the
                     following tissue(s): Primary Lung Epithelial Cells The
                     library was constructed according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. First strand
                     cDNA synthesis was primed with an oligo-dT primer
                     containing a Not I site.  Double stranded cDNA was ligated
                     to an EcoR I adaptor, digested with Not I, and cloned
                     directionally into pT7T3-Pac vector. The oligonucleotide
                     used to prime the synthesis of first-strand cDNA contains
                     a library tag sequence that is located between the Not I
                     site and the (dT)18 tail. The sequence tag for this
                     library is GGCTGTAGGC.
                     TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
                     TAG_LIB=UI-CF-DU1
                     TAG_SEQ=GGCTGTAGGC"
ORIGIN

  Query Match           15.2%;  Score 304.2;  DB 5;  Length 696;
  Best Local Similarity   73.7%;  Pred. No. 3.5e-38;
  Matches  505;  Conservative    0;  Mismatches  145;  Indels   35;  Gaps    8;

Qy       1166 TACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGT 1225
              ||| ||||||| |||||  |||||| |||||| |||||||| || || | ||  ||||||

Db         18 TACTTTAATTTAAAAATATTTTATTGCTAAAAGATGCTAACGATAATCTAAGTCTTCAGT 77

Qy       1226 GAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGGTCA- 1284
              ||||       || |||||||  ||  |||||  || | || |||  | |   |  | |

Db         78 GAGTCT-----TTGTTGCTGGAAGACAGTCTTGCCTCACTGTTGATGGCTGCTGAATGAT 132

Qy       1285 ----------GGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAG 1329
                        | ||||||||  |||| ||||||  | |    ||||||  ||     ||||||

Db        133 CGGTGTGGTGGTTGCTGAAGGTTAGTGTGGCTCTAGAGATTTCTTTAAATGAGACAACAA 192

Qy       1330 TGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGAT 1389
              ||||  |||||  ||||  |||||||||||||  |||||||| ||| |||  |||| |

Db        193 GGAAGGCTGCAACATCAACTGACTCTTCCTTTC-CGAAAGATTTATCTGTAGCATGTGTT 251

Qy       1390 GCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTC 1448
              ||| || ||||||  |||||  ||||||  ||||  |  || |||||||  | ||||||||||

Db        252 GCTGTTGGATAGCCTTTTACCCACAGTGGAACATTCTTCAAAATTGAAGTCAATCCTCTC 311

Qy       1449 AAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCAT 1508
              ||||  |||||     |||  |  |||||||  || |||||||||  | |||  ||||||||||

Db        312 AAACTCTGCTGCTGTTTTATCAACTAAGTTTATGTAATATTCTAAGTCCTTTGTTGTCAT 371

Qy       1509 TTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGA 1568
              |||||||||  ||||||  |||||||||||||||  ||||||||||        |  |

Db        372 TTCAACAATGGTCACAGGATCTTCACCAGGAGTAGTTTCCATCTCAAGAAACCACTT--- 428

Qy       1569 ATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCA 1628
                |  | ||||  |  || |||||  || |||||||  ||||||  ||||||||||||||||||||

Db        429 -TTGTGGCTCTTATATTAGAAGCAACTCCTCATTTGTTCACGTTTTATCATGAGATTGCA 487

Qy       1629 GCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTC 1688
              |||||  ||||||  |  |  ||||       |  |||||  |||  ||  ||||  |||||||  ||||

Db        488 GCAATTCAGTCACATATGCAGG-----TTCCACTTCTAAGTCTAGTTATCTTGCTATTTC 542

Qy       1689 TACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAG 1748
              ||||||||||  |||  ||||  ||||  ||||||||||||||  |||  |  ||||||||  ||

Db        543 CACCACATCTGCAGTTACTTCGTCCACTGAAGTCTTGAACCCCTCAAGATCATCCATCAG 602

Qy       1749 GGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC---CTCCCATGA 1805
              |||||   ||  |  |||||  ||  ||||||||||  ||  ||||||||||   |||  |||||

Db        603 GGTTGAGGTCAATGTCTTCTGAACTCCTGTTAATGTTGATATTTTGACCTTCTCNCATGA 662

Qy       1806 ATCATGAATGTTCTTAATGGCACCT 1830
              ||||   |||| ||||||||||  ||

Db        663 ATCACANATGTCCTTAATGGCATCT 687


RESULT 44
DN100706
LOCUS       DN100706                849 bp    mRNA    linear   EST 14-FEB-2005
DEFINITION  1000083 MARC 4PIG Sus scrofa cDNA 3', mRNA sequence.
ACCESSION   DN100706
VERSION     DN100706.1  GI:59773001
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
            Sus.
REFERENCE   1  (bases 1 to 849)
  AUTHORS   Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
            Nonneman,D.J., Wray,J.E. and Keele,J.W.
  TITLE     Porcine EST collection using a normalized library constructed from
            embryos representing early developmental stages
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
```

```
          cross_match v0.990329.
          Plate: TMW8069  row: P  column: 22
          Seq primer: TAGAAGGCACAGTCGAGG.
FEATURES             Location/Qualifiers
     source          1. .849
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 4PIG"
                     /note="Vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI;
                     Library made with combined RNA from day-10, day-13,
                     day-15, day-25, and day-30 whole embryos."
ORIGIN

  Query Match             15.2%;  Score 304;  DB 8;  Length 849;
  Best Local Similarity   69.1%;  Pred. No. 3.6e-38;
  Matches  612;  Conservative    0;  Mismatches  215;  Indels   59;  Gaps   12;

Qy       1121 TGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAA 1180
              | | |||| ||   | |  |     ||  ||       ||   ||||||||| |||  |||
Db             3 TTTGGTCTTTTGTCTTTTGAGGGCCGTACTGCGGCCCAATGTACATACCTTCATTAAAAA 62

Qy       1181 ATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTT 1240
              || ||||||| || |||| ||||||| ||||| | ||  || ||  |||||| |||||||
Db            63 ATACTTTATTGCT-AAAATTGCTAACCATCATCTAAGTCTTTAGCAAGTTGTGATCTTTT 121

Qy       1241 TGCTGGTGGAAGGTCT---TTTCTTATTGATGACTGAT--------CGGGGGTCAGGTGC 1289
              |  |||| || |||||   |||  ||||||||| || |         ||| |   | ||
Db           122 TATTGGTAGAGGGTCTTGCTTTGATATTGATGATTGCTGGCCAACCAGGGTGGTGGCTGT 181

Qy       1290 TGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA--ACAACAGTGAAGATTGCAATATCAG 1347
              |||    | || ||| | ||||| |||||||||  | |||| | |||  | | | ||
Db           182 TGACAGCTGGGATGGTTATGGCAATTTCTTAAAATAGAACAATAAAGTATACCACATGGC 241

Qy       1348 TTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTT 1407
              ||| ||  |||||||||||  |||||||| |||  ||  |||||  |||| |||||||
Db           242 TTGGCTACTCCTTTCATGAACAATTTCTCTGTAGCATGCAATGCTACTTGACAGCATTTA 301

Qy       1408 ATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACCCTGCTCTGCTTTAA 1467
                   | || ||  ||||| |  ||| | |||| ||| ||||| ||||||  |
Db           302 CCAACAGCAGAACTTTTTTCAAAAATGGAGTCACTTCTCTTAAATCCTGCACTGCTTCAT 361

Qy       1468 CAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTG 1527
              ||||     || |||||||||||| | ||| ||||||| |||||||| || |||||
Db           362 CAAC-----TTCATATAATATTCTAAGTCCTTTGTTGTTATTTCAACCATCTTCAC---- 412

Qy       1528 TCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAG 1587
                    | ||| ||| ||||||        |  |    |||||||| || ||||||
Db           413 ----------CAATAGGTTCTATCTCAAGAAACCATTT----TCTTTGCTTATACATAAG 458

Qy       1588 AAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTC 1647
              ||| | ||||| |||  || |||||||||||||||||||||||| ||||||| | |||
Db           459 AAGCAGTTCCTAATCCATTTAAGTTTTATCATGAGATTGCAGCAATTCAGTCATATATTC 518

Qy       1648 AGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCT 1707
              | |      |||| || ||||||   ||| | |||| |||||  ||||||||   |  ||
Db           519 ATG-----TTTCATTTCTAATTCTGATTCCCCTGCTATTTCTTTCACATCTGCAATCACT 573

Qy       1708 TCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTC 1767
              | ||||| |||||||||||| | ||         ||  |||||||  || | |||| |||
Db           574 TACTCCAGTGAAGTCTTGAAACCCT--------TCTGTGAGGGTTACAACCAACTTTTTC 625

Qy       1768 CAAATTCCTGTTAATATTTATATTTTGACC---TCCCATGAATCATGAATGTTCTTAATG 1824
              ||||||||| |||||||| ||||||||||   |||||||||  |||||||| ||||||
Db           626 CAAATTCCTATTAATATTGATATTTTGACCTCGTCCCATGAATAGTGAATGTTGTTAATG 685

Qy       1825 GCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCC 1884
              || || ||||||| ||| |||||| |||| |||||||   |||| | ||||||||||
Db           686 ACATCTAGAATGGTAAATTCTTTCCCAAAGATTTTCAAACAACTTCGCCCAGATCCAT-- 743

Qy       1885 ATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAAT 1944
                    |||||||  ||||||| ||  ||| ||  ||||| |  |||||| ||||||||||
Db           744 ---CAGAGGAATCACTTTCTATTGCAGCTACTGCCTT-TCAAATGTAATTCTTCAATAAA 799

Qy       1945 AAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAAT 1990
                 |||   | ||| ||||||| |||||||||||| |   ||| | |||
Db           800 AAGATCTAAAAATTGAAATGACTCCTTGATCCCTGGACTGTAGAAT 845


RESULT 45
CA308451/c
LOCUS       CA308451                620 bp    mRNA    linear   EST 05-AUG-2004
DEFINITION  UI-H-FT1-bhz-b-08-0-UI.s1 NCI_CGAP_FT1 Homo sapiens cDNA clone
            UI-H-FT1-bhz-b-08-0-UI 3', mRNA sequence.
ACCESSION   CA308451
VERSION     CA308451.1  GI:24471505
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Gary W. Hunninghake, U of I
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
            sequence: 18-620, >TIGGER1#DNA/MER2_type
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FT1-bhz-b-08-0-UI"
                     /tissue_type="Alveolar Macrophage"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_FT1"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_FT1 is a normalized cDNA library constructed from
                     a pool of 81 RNA samples from Alveolar Macrophages
                     challenged with different treatments. The mRNA samples
                     were a mixture of these conditions (times refer to
                     incubations following isolation by bronchoalveolar lavage)
                     (some normal donor macrophages were cultured in some of
                     the conditions, other donor macrophages in different
                     conditions). The mRNA samples were pooled for library
                     construction. Control 0 hours; control 3 hours; control 24
                     hours; LPS 100 ng/ml, 3 hours; LPS 100 ng/ml, 24 hours;
                     PMA 10 ng/ml, 3 hours; PMA 10 ng/ml, 24 hours; Klebsiella
                     moi 10, 3 hours; Klebsiella moi 10, 24 hours; Staph aureus
                     moi 10, 3 hours; Staph aureus moi 10, 24 hours; Adenoviral
                     vector (Ad5 CMV eGFP), moi 500, 3 hours; Adenoviral vector
                     (Ad5 CMV eGFP), moi 500, 24 hours; wt adenovirus moi 500,
                     3 hours; wt adenovirus moi 500, 24 hours; Ad vector + LPS
                     3 hours; Ad vector + LPS 24 hours; wt adenovirus + LPS 3
                     hours; wt adenovirus + LPS 24 hours. The library was
                     normalized according to Bonaldo, Lennon and Soares, Genome
                     Research, 6:791-806, 1996. First strand cDNA synthesis was
                     primed with an oligo-dT primer containing a Not I site.
                     Double stranded cDNA was ligated to an EcoR I adaptor,
                     digested with Not I, and cloned directionally into
```

pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GGCCATGCCG. The tissue was provided by Dr. Gary W. Hunninghake of the University of Iowa.
TAG_TISSUE=Human Lung Alveolar Macrophage
TAG_LIB=UI-H-FT1
TAG_SEQ=GGCCATGCCG"

ORIGIN

Query Match 15.2%; Score 303; DB 6; Length 620;
Best Local Similarity 73.4%; Pred. No. 5.6e-38;
Matches 467; Conservative 0; Mismatches 136; Indels 33; Gaps 5;

```
Qy       1330 TGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGAT 1389
             ||||| |||| | |||| | |||| |  ||||||  |||||||||||| ||  |  ||
Db        620 TGAAGCTTGCCACATCAATGGACTTTCTCTTTCAGAAAAGATTTCTCTGTAAAGCATGTG 561

Qy       1390 GCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCA 1449
                | ||        || | ||||    |  |  ||  ||| |||  ||||||||||||
Db        560 ATGATGTGC------ATTTTACACACAGAACTTCTTTCAAAATTTGAGTCAATCCTCTCA 507

Qy       1450 AACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATT 1509
             |||||||||     || |    |||| || || |||||||   ||||| ||||||||||||
Db        506 AACCCTGCTGCTGCTTTATTGACTAAATTGATGTAATATTACAAATCCTTTGTTGTCATT 447

Qy       1510 TCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAA 1569
             |||||||| || ||||  |||||||||||||||||||||||||||     | |  |
Db        446 TCAACAATGTTGACAGCATCTTCACCAGGAGTAGATTCCATCTCAAGAAATCACTT---- 391

Qy       1570 TCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAG 1629
             ||||  |||||||||||||| || |||||||| ||  ||||||||| ||||||||||||
Db        390 TCTTCACTCATCCATAAGAAGCAACTCCTCATCAGTGAAAGTTTTATAATGAGATTGCAG 331

Qy       1630 CAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCT 1689
             ||||  |||                  || ||||| |||||| |||||  ||||| |||||
Db        330 CAATTCAG----------------GCTCCACTTCTAATTCTAGTTCATTTGCTATTTCT 288

Qy       1690 ACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGG 1749
             |||||||  |  ||| |||||||||| |||||||||| || || | || ||||||| || |
Db        287 ACCACATAGGCAGTTACTTCCTCCACTGAAGTCTTGGACTTCACAAAATCATCCACGAAG 228

Qy       1750 GTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCCCATGAATCA 1809
              ||||||||  ||| |||||||| |||| ||| |  | ||||||||| || ||||| ||||
Db        227 TTTGGAATCAGCTTTTTCCAAACTCCTATTAGTGCTGATATTTTGA-CTTCCATGCATCA 169

Qy       1810 TGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTT 1869
             |||||||||||||||| | || ||||||||||||||||||| |||| |||||||||  |||
Db        168 TGAATGTTCTTAATGGTATCTAGAATGGTGAATCCTTTCCAGAAGGCTTTCAATTTGCTT 109

Qy       1870 AGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATG 1929
              | ||||  ||||      |||||||  ||   || ||| ||| ||||||||||| |||
Db        108 TGCCCAGGACCAT-----CAGAGGAATCATNATCTATGGCAGCTATAGCCTTATGAAATT 54

Qy       1930 TATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTA 1965
             |||||||| ||||| | | |||||||||  | ||||
Db         53 TATTTCTTAAATAAAAGGACTTGAAAGTCAAGATTA 18
```

Search completed: January 22, 2006, 21:08:38
Job time : 5438.83 secs

This Page Intentionally Left Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 01:00:32 ; Search time 6805.08 Seconds
(without alignments)
16706.187 Million cell updates/sec

Title: US-09-728-552A-3_COPY_1_2000
Perfect score: 2000
Sequence: 1 gaattctcctgcctcagcct.....tctgcaaaatagatgttgtg 2000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2000	100.0	40917	6	AX033911	AX033911 Sequence
2	2000	100.0	80622	8	AF222855	AF222855 Homo sapi
3	1966	98.3	112442	8	AL355340	AL355340 Human DNA
c 4	1966	98.3	176432	14	AL391648	AL391648 Homo sapi
5	1611.4	80.6	41008	6	AX033912	AX033912 Sequence
6	1592.8	79.6	69058	8	AF222856	AF222856 Homo sapi
7	1592.8	79.6	80202	8	AF222854	AF222854 Homo sapi
8	1575.2	78.8	80155	8	AF042484	AF042484 Homo sapi
c 9	636.8	31.8	2418	6	AR612649	AR612649 Sequence
c 10	634	31.7	2418	8	HSU49973	U49973 Human Tigge
11	526.6	26.3	100020	8	AC092165	AC092165 Homo sapi
12	526.6	26.3	158276	14	AC012033	AC012033 Homo sapi
13	526.6	26.3	179816	14	AC074188	AC074188 Homo sapi
14	526.6	26.3	210301	8	AF307337	AF307337 Homo sapi
c 15	523.4	26.2	2641	8	BC035143	BC035143 Homo sapi
c 16	522.2	26.1	60169	8	AL596257	AL596257 Human DNA
17	518.8	25.9	51343	8	AC108017	AC108017 Homo sapi
18	516.8	25.8	135818	14	AC158447	AC158447 Loxodonta
19	516.8	25.8	141510	14	AC157635	AC157635 Loxodonta
20	516	25.8	3895	6	AX833395	AX833395 Sequence
21	516	25.8	3895	8	AK095077	AK095077 Homo sapi
c 22	515.2	25.8	115958	8	AC004736	AC004736 Human Chr
23	515.2	25.8	115971	8	AC124056	AC124056 Homo sapi
c 24	515.2	25.8	145713	8	AC124301	AC124301 Homo sapi
c 25	514.4	25.7	178098	8	HS168L15	AL022069 Human DNA
26	511.2	25.6	73991	8	AL357332	AL357332 Human DNA
c 27	511.2	25.6	114056	8	AL590674	AL590674 Human DNA
c 28	511.2	25.6	128479	8	AL596266	AL596266 Human DNA
c 29	511	25.6	99025	8	HSDJ45N11	AL132996 Human DNA
30	511	25.6	172061	14	AC027156	AC027156 Homo sapi
31	510.4	25.5	143747	8	HS370M22	Z82206 Human DNA s
c 32	508.2	25.4	173691	8	AC022101	AC022101 Homo sapi
c 33	507.6	25.4	98517	8	HS366N23	AL021331 Human DNA
c 34	506.2	25.3	110907	8	AC084703	AC084703 Homo sapi
c 35	506.2	25.3	183794	8	AC013752	AC013752 Homo sapi
c 36	505.2	25.3	156320	8	AC022028	AC022028 Homo sapi
c 37	505.2	25.3	159771	8	AL139142	AL139142 Human DNA
38	505.2	25.3	162163	14	AC021635	AC021635 Homo sapi
c 39	505.2	25.3	169424	14	AL929203	AL929203 Homo sapi
40	505.2	25.3	188759	6	CQ870480	CQ870480 Sequence
41	504	25.2	175210	8	CNS07EG5	AL590328 Human chr
42	504	25.2	200853	8	CNS01DSQ	AL121839 Human chr
43	503.4	25.2	68048	8	AC004764	AC004764 Homo sapi
c 44	503.4	25.2	126034	8	AC008670	AC008670 Homo sapi
45	503.4	25.2	178942	8	AL357520	AL357520 Human DNA
c 46	503.4	25.2	245383	14	AL161653	AL161653 Homo sapi
47	501.8	25.1	136961	8	AC136290	AC136290 Homo sapi
c 48	501.8	25.1	140549	8	AP006192	AP006192 Homo sapi
c 49	501.8	25.1	163908	8	AC005411	AC005411 Homo sapi
c 50	501.8	25.1	200000	8	AP000493	AP000493 Homo sapi
c 51	501	25.1	179590	8	AC115986	AC115986 Homo sapi
52	501	25.1	180005	14	AC013532	AC013532 Homo sapi
c 53	501	25.1	180917	8	AC018517	AC018517 Homo sapi
c 54	501	25.1	185802	14	AC073843	AC073843 Homo sapi
c 55	500.6	25.0	170333	14	AL139019	AL139019 Homo sapi
56	500.6	25.0	192589	8	AC093157	AC093157 Homo sapi
57	500.4	25.0	5376	8	HSM803941	AL832631 Homo sapi
58	500.2	25.0	138350	14	AC010801	AC010801 Homo sapi
c 59	500	25.0	139273	8	AL596275	AL596275 Human DNA
60	500	25.0	187999	14	AC026928	AC026928 Homo sapi
c 61	496	24.8	213721	8	HS172B20	AL022319 Human DNA
62	495.2	24.8	211662	8	AL391825	AL391825 Human DNA
c 63	494.6	24.7	125565	8	AL359762	AL359762 Human DNA
c 64	493	24.6	179844	8	AC009265	AC009265 Homo sapi
c 65	492.8	24.6	103488	8	AL805915	AL805915 Human DNA
66	492.8	24.6	153771	14	AC037463	AC037463 Homo sapi
67	492.8	24.6	160396	8	AL592063	AL592063 Human DNA
c 68	492.8	24.6	168460	14	AC011585	AC011585 Homo sapi
c 69	492.8	24.6	182912	8	AC068060	AC068060 Homo sapi
70	492.8	24.6	182913	8	AC068556	AC068556 Homo sapi
71	492.8	24.6	201133	14	AL807235	AL807235 Homo sapi
c 72	492.6	24.6	147760	8	AC011846	AC011846 Homo sapi
c 73	490.8	24.5	202874	8	AL389888	AL389888 Human DNA
74	490.6	24.5	140100	8	AC016778	AC016778 Homo sapi
75	490	24.5	168153	8	AF127577	AF127577 Homo sapi
76	490	24.5	183436	8	AC012499	AC012499 Homo sapi
77	490	24.5	340000	8	HS21C008	AL163208 Homo sapi
78	489.8	24.5	214175	8	CNS01DRT	AL118557 Human chr
79	489.4	24.5	179902	8	CNS01DRN	AL117672 Human chr
c 80	489	24.4	195981	8	AC007652	AC007652 Homo sapi
81	488	24.4	127683	8	AC003683	AC003683 Homo sapi
c 82	486.8	24.3	175579	8	AC007768	AC007768 Homo sapi
c 83	486.8	24.3	199702	14	AC087178	AC087178 Homo sapi
84	486.4	24.3	137520	4	ECA459965	AJ459965 Equus cab
c 85	486.4	24.3	188165	8	CNS00003	AL049781 Human chr
86	485.4	24.3	163106	14	AP002375	AP002375 Homo sapi
c 87	485.4	24.3	169755	14	AP001199	AP001199 Homo sapi
c 88	485.4	24.3	190066	8	AC010127	AC010127 Homo sapi
89	485.4	24.3	191194	14	AC021673	AC021673 Homo sapi
c 90	485.2	24.3	194635	8	AC012087	AC012087 Homo sapi
c 91	484.6	24.2	196799	8	AC146074	AC146074 Pan trogl

c	92	484	24.2	172345	8	AC084859 AC084859 Homo sapi
c	93	483.8	24.2	147394	8	HS353H6 AL022577 Human DNA
c	94	483.6	24.2	153364	14	AC009787 AC009787 Homo sapi
	95	483.6	24.2	156545	8	AC112217 AC112217 Homo sapi
c	96	483.4	24.2	94817	8	HS981L23 AL031686 Human DNA
	97	483	24.1	138391	14	AL357733 AL357733 Homo sapi
	98	483	24.1	209105	14	AC161886 AC161886 Pan trogl
	99	482.6	24.1	177472	8	AC008786 AC008786 Homo sapi
c	100	482.4	24.1	144048	8	AC013762 AC013762 Homo sapi

ALIGNMENTS

RESULT 1
AX033911
LOCUS AX033911 40917 bp DNA linear PAT 21-SEP-2000
DEFINITION Sequence 3 from Patent WO9851790.
ACCESSION AX033911
VERSION AX033911.1 GI:10280479
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified
unclassified sequences.
REFERENCE 1
AUTHORS Cancilla,M.R., Choo,K.H. and Du,S.D.
TITLE A novel nucleic acid molecule
JOURNAL Patent: WO 9851790-A 3 19-NOV-1998;
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES Location/Qualifiers
source 1. .40917
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN

Query Match 100.0%; Score 2000; DB 6; Length 40917;
Best Local Similarity 100.0%; Pred. No. 1.7e-300;
Matches 2000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60

Qy         61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120

Qy        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180

Qy        181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240

Qy        241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300

Qy        301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 360

Qy        361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420

Qy        421 TATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480

Qy        481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540

Qy        541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600

Qy        601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660

Qy        661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720

Qy        721 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780

Qy        781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840

Qy        841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATGTCTATAATACATT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATGTCTATAATACATT 900

Qy        901 AGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCT 960

Qy        961 GTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCAC 1020

Qy       1021 CACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTT 1080

Qy       1081 TGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA 1140

Qy       1141 ATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAAT 1200

Qy       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260

Qy       1261 TTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320

Qy       1321 AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCA 1440

Qy       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500

Qy       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560

Qy       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620

Qy       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680

Qy       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740

Qy       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800

Qy       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860

Qy       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920

Qy       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980

Qy       1981 TCTGCAAAATAGATGTTGTG 2000
                ||||||||||||||||||||
Db       1981 TCTGCAAAATAGATGTTGTG 2000
```

```
RESULT 2
AF222855
LOCUS       AF222855               80622 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222855
VERSION     AF222855.1  GI:9246845
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80622)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 80622)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Genomic sequence from human 10q25.2, clonelib=HC
FEATURES             Location/Qualifiers
     source          1. .80622
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="HC"
     repeat_region   complement(3. .176)
                     /rpt_family="Alu"
     repeat_region   complement(979. .3564)
                     /rpt_family="Tigger1"
     repeat_region   2967. .3094
                     /rpt_family="Alu"
     repeat_region   3664. .3934
                     /rpt_family="Alu"
     repeat_region   complement(4920. .5181)
                     /rpt_family="Alu"
     repeat_region   complement(6287. .6579)
                     /rpt_family="Alu"
     repeat_region   6625. .6966
                     /rpt_family="MLT2B2"
     repeat_region   complement(9215. .9490)
                     /rpt_family="Alu"
     repeat_region   complement(11280. .11565)
                     /rpt_family="Alu"
     repeat_region   complement(12428. .12675)
                     /rpt_family="Alu"
     repeat_region   complement(14304. .14583)
                     /rpt_family="Alu"
     repeat_region   17503. .17794
                     /rpt_family="Alu"
     repeat_region   complement(20250. .20533)
                     /rpt_family="Alu"
     repeat_region   complement(23031. .23323)
                     /rpt_family="Alu"
     repeat_region   complement(24301. .24479)
                     /rpt_family="MIR"
     repeat_region   24951. .25250
                     /rpt_family="Alu"
     repeat_region   complement(25987. .26342)
                     /rpt_family="THE1"
     repeat_region   27097. .27335
                     /rpt_family="MER2"
     repeat_region   27538. .27814
                     /rpt_family="Alu"
     repeat_region   27891. .28161
                     /rpt_family="Alu"
     repeat_region   29063. .29334
                     /rpt_family="Alu"
     repeat_region   complement(31025. .31093)
                     /rpt_family="MLT1"
     repeat_region   complement(31817. .32352)
                     /rpt_family="MER1"
     repeat_region   complement(32489. .32607)
                     /rpt_family="MLT1"
     repeat_region   33039. .33316
                     /rpt_family="Alu"
     repeat_region   33469. .33760
                     /rpt_family="Alu"
     repeat_region   complement(34527. .34590)
                     /rpt_family="MIR"
     repeat_region   34689. .34995
                     /rpt_family="Alu"
     repeat_region   complement(36349. .36437)
                     /rpt_family="MIR"
     repeat_region   complement(38796. .39079)
                     /rpt_family="Alu"
     repeat_region   39441. .39693
                     /rpt_family="Alu"
     repeat_region   complement(43176. .43237)
                     /rpt_family="MIR"
     repeat_region   complement(45203. .45478)
                     /rpt_family="Alu"
     repeat_region   45892. .46025
                     /rpt_family="MIR"
     repeat_region   complement(46618. .46851)
                     /rpt_family="Alu"
     repeat_region   48623. .48939
                     /rpt_family="Alu"
     repeat_region   complement(49048. .50698)
                     /rpt_family="L1"
     repeat_region   complement(50710. .50972)
                     /rpt_family="Alu"
     repeat_region   complement(51031. .51192)
                     /rpt_family="L1"
     repeat_region   complement(51450. .51739)
```

```
                     /rpt_family="Alu"
     repeat_region   complement(53270. .53562)
                     /rpt_family="Alu"
     repeat_region   54261. .54808
                     /rpt_family="LTR12"
     repeat_region   59460. .60054
                     /rpt_family="LTR12"
     repeat_region   60349. .60633
                     /rpt_family="Alu"
     repeat_region   complement(60669. .60801)
                     /rpt_family="MIR"
     repeat_region   complement(61684. .61852)
                     /rpt_family="MER5"
     repeat_region   63466. .63753
                     /rpt_family="Alu"
     repeat_region   64063. .64228
                     /rpt_family="MER5"
     repeat_region   64446. .64521
                     /rpt_family="MIR"
     repeat_region   65794. .65883
                     /rpt_family="MER5"
     repeat_region   66772. .66951
                     /rpt_family="MER20"
     repeat_region   complement(67895. .68416)
                     /rpt_family="MER41"
     repeat_region   complement(73192. .73470)
                     /rpt_family="Alu"
     repeat_region   74086. .74367
                     /rpt_family="Alu"
     repeat_region   75916. .76183
                     /rpt_family="Alu"
     repeat_region   complement(77771. .78059)
                     /rpt_family="Alu"
ORIGIN

  Query Match             100.0%;  Score 2000;  DB 8;  Length 80622;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-300;
  Matches 2000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60

Qy         61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120

Qy        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180

Qy        181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240

Qy        241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300

Qy        301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 360

Qy        361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420

Qy        421 TATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480

Qy        481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540

Qy        541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600

Qy        601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660

Qy        661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720

Qy        721 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780

Qy        781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840

Qy        841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATGTCTATAATACATT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATGTCTATAATACATT 900

Qy        901 AGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCT 960

Qy        961 GTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCAC 1020

Qy       1021 CACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTT 1080

Qy       1081 TGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA 1140

Qy       1141 ATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAAT 1200

Qy       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260

Qy       1261 TTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320

Qy       1321 AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCA 1440

Qy       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500

Qy       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560

Qy       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
```

```
Qy       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680

Qy       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740

Qy       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800

Qy       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860

Qy       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920

Qy       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980

Qy       1981 TCTGCAAAATAGATGTTGTG 2000
             ||||||||||||||||||||
Db       1981 TCTGCAAAATAGATGTTGTG 2000
```

RESULT 3
AL355340
LOCUS AL355340 112442 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-383C6 on chromosome 10 Contains the TRUB1 gene for TruB pseudouridine (psi) synthase homolog 1 (E. coli), a novel gene and a CpG island, complete sequence.
ACCESSION AL355340
VERSION AL355340.17 GI:21436504
KEYWORDS HTG; TRUB1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 112442)
AUTHORS Leongamornlert,D.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 17, 2002 this sequence version replaced gi:16416169.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr10
RP11-383C6 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vega@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES Location/Qualifiers
source 1. .112442
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIB753C06383"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-383C6"
/clone_lib="RPCI-11.2"
misc_feature 2000
/note="Clone_right_end: RP11-106M7"
gene join(24099. .24445,28551. .28649,37000. .37055,45632. .45713, 56274. .56346,58041. .58180,60235. .60291,61029. .63577)
/gene="TRUB1"
/locus_tag="RP11-383C6.1-001"
mRNA join(24099. .24445,28551. .28649,37000. .37055,45632. .45713, 56274. .56346,58041. .58180,60235. .60291,61029. .63577)
/gene="TRUB1"
/locus_tag="RP11-383C6.1-001"
/product="TruB pseudouridine (psi) synthase homolog 1 (E. coli)"
/note="match: ESTs: AI208778.1 BQ775335.1
match: cDNAs: AF448144.1 BC030601.1"
CDS join(24160. .24445,28551. .28649,37000. .37055,45632. .45713, 56274. .56346,58041. .58180,60235. .60291,61029. .61285)
/gene="TRUB1"
/locus_tag="RP11-383C6.1-001"
/standard_name="OTTHUMP00000020554"
/note="match: proteins: Q8WWH5 Q9D331"
/codon_start=1
/product="TruB pseudouridine (psi) synthase homolog 1 (E. coli)"
/protein_id="CAH73219.1"
/db_xref="GI:55663184"
/db_xref="Genew:16060"
/db_xref="GOA:Q8WWH5"
/db_xref="InterPro:IPR002501"
/db_xref="InterPro:IPR004510"
/db_xref="UniProt/TrEMBL:Q8WWH5"
/translation="MAASEAAVVSSPSLKTDTSPVLETAGTVAAMAATPSARAAAAVV AAAARTGSEARVSKAALATKLLSLSGVFAVHKPKGPTSAELLNRLKEKLLAEAGMPSP EWTKRKKQTLKIGHGGTLDSAARGVLVVGIGSGTKMLTSMLSGSKRYTAIGELGKATD TLDSTGRVTEEKPYDKITQEDIEGILQKFTGNIMQVPPLYSALKKDGQRLSTLMKRGE VVEAKPARPVTVYSISLQKFQPPFFTLDVECGGGFYIRSLVSDIGKELSSCANVLELT RTKQGPFTLEEHALPEDKWTIDDIAQSLEHCSSLFPAELALKKSKPESNEQVLSCEYI TLNEPKREDDVIKTC"
gene join(24292. .24445,28551. .28649,37000. .37055,45632. .45906)
/gene="TRUB1"
/locus_tag="RP11-383C6.1-002"
mRNA join(24292. .24445,28551. .28649,37000. .37055,45632. .45906)
/gene="TRUB1"
/locus_tag="RP11-383C6.1-002"
/product="TruB pseudouridine (psi) synthase homolog 1 (E. coli)"
/note="match: ESTs: AI418848.1 AI422325.1 AI640751.1 BE858241.1"
polyA_signal 61455. .61460
/gene="TRUB1"
/locus_tag="RP11-383C6.1-001"
polyA_site 61479
/gene="TRUB1"
/locus_tag="RP11-383C6.1-001"
polyA_signal 63558. .63563
/gene="TRUB1"
/locus_tag="RP11-383C6.1-001"
polyA_site 63582
gene join(80563. .80691,81349. .81444,82381. .82499)
/locus_tag="RP11-383C6.2-001"
mRNA join(80563. .80691,81349. .81444,82381. .82499)

```
                     /locus_tag="RP11-383C6.2-001"
                     /product="putative novel transcript"
                     /note="match: ESTs: AA774571.1"
     misc_feature    110443
                     /note="Clone_left_end: RP11-338L11"
ORIGIN

  Query Match             98.3%;  Score 1966;  DB 8;  Length 112442;
  Best Local Similarity   99.9%;  Pred. No. 2.1e-295;
  Matches 1999;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68556 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 68615

Qy         61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68616 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 68675

Qy        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68676 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 68735

Qy        181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68736 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 68795

Qy        241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68796 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 68855

Qy        301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 360
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      68856 GTCATTATGAATGTCTGTATTCATT-CCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 68914

Qy        361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68915 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 68974

Qy        421 TATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68975 TATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 69034

Qy        481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69035 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 69094

Qy        541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69095 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 69154

Qy        601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69155 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 69214

Qy        661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69215 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 69274

Qy        721 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69275 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 69334

Qy        781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69335 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 69394

Qy        841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTA-TGTCTATAATACAT 899
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      69395 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATTGTCTATAATACAT 69454

Qy        900 TAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTC 959
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      69455 TAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTC 69514

Qy        960 TGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCA 1019
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69515 TGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCA 69574

Qy       1020 CCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTT 1079
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69575 CCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTT 69634

Qy       1080 TTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGC 1139
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69635 TTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGC 69694

Qy       1140 AATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAA 1199
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69695 AATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAA 69754

Qy       1200 TGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTT 1259
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69755 TGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTT 69814

Qy       1260 CTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT 1319
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69815 CTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT 69874

Qy       1320 AAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69875 AAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 69934

Qy       1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-T 1438
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      69935 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGAGT 69994

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69995 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 70054

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70055 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 70114

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70115 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 70174

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70175 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 70234

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70235 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 70294

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70295 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 70354

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70355 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 70414

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70415 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 70474

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70475 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 70534

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000
              |||||||||||||||||||||||
Db      70535 TTTCTGCAAAATAGATGTTGTG 70556
```

```
RESULT 4
AL391648/c
LOCUS       AL391648              176432 bp    DNA     linear   HTG 18-JUL-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-203A7, 9 unordered pieces.
ACCESSION   AL391648
VERSION     AL391648.12  GI:14669309
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 10, 2001 this sequence version replaced gi:14596357.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA203A7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 174005 bases at least Q40
            Consensus quality: 174806 bases at least Q30
            Consensus quality: 175273 bases at least Q20
            Insert size: 175632; sum-of-contigs
            Insert size: 155595; 14.1% error; agarose-fp
            Quality coverage: 6.85x in Q20 bases; sum-of-contigs Quality
            coverage: 8.13x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    23099: contig of 23099 bp in length
            *   23100    23199: gap of 100 bp
            *   23200    38049: contig of 14850 bp in length
            *   38050    38149: gap of 100 bp
            *   38150    65430: contig of 27281 bp in length
            *   65431    65530: gap of 100 bp
            *   65531   100606: contig of 35076 bp in length
            *  100607   100706: gap of 100 bp
            *  100707   156710: contig of 56004 bp in length
            *  156711   156810: gap of 100 bp
            *  156811   159233: contig of 2423 bp in length
            *  159234   159333: gap of 100 bp
            *  159334   171732: contig of 12399 bp in length
            *  171733   171832: gap of 100 bp
            *  171833   174095: contig of 2263 bp in length
            *  174096   174195: gap of 100 bp
            *  174196   176432: contig of 2237 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176432
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-203A7"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .23099
                     /note="assembly_fragment:01553
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    23200. .38049
                     /note="assembly_fragment:00834
                     fragment_chain:1"
     misc_feature    38150. .65430
                     /note="assembly_fragment:02112
                     fragment_chain:1"
     misc_feature    65531. .100606
                     /note="assembly_fragment:01440
                     fragment_chain:1"
     misc_feature    100707. .156710
                     /note="assembly_fragment:01209
                     fragment_chain:1"
     misc_feature    156811. .159233
                     /note="assembly_fragment:00074"
     misc_feature    159334. .171732
                     /note="assembly_fragment:01455"
     misc_feature    171833. .174095
                     /note="assembly_fragment:01550"
     misc_feature    174196. .176432
                     /note="assembly_fragment:03092"
ORIGIN

  Query Match             98.3%;  Score 1966;  DB 14;  Length 176432;
  Best Local Similarity   99.9%;  Pred. No. 1.7e-295;
  Matches 1999;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44960 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 44901

Qy         61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44900 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 44841

Qy        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44840 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 44781

Qy        181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44780 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 44721

Qy        241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44720 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 44661

Qy        301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 360
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      44660 GTCATTATGAATGTCTGTATTCATT-CCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 44602

Qy        361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44601 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 44542

Qy        421 TATTTTTCCTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44541 TATTTTTCCTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 44482

Qy        481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 44422

Qy        541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44421 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 44362

Qy        601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      44361 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 44302

Qy        661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44301 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 44242

Qy        721 TTTAGTTATACCTCTTAAAATTTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44241 TTTAGTTATACCTCTTAAAATTTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 44182

Qy        781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44181 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 44122

Qy        841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTA-TGTCTATAATACAT 899
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      44121 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATTGTCTATAATACAT 44062

Qy        900 TAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTC 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44061 TAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTC 44002

Qy        960 TGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44001 TGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCA 43942

Qy       1020 CCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTT 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43941 CCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTT 43882

Qy       1080 TTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43881 TTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGC 43822

Qy       1140 AATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAA 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43821 AATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAA 43762

Qy       1200 TGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTT 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43761 TGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTT 43702

Qy       1260 CTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43701 CTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT 43642

Qy       1320 AAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43641 AAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 43582

Qy       1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-T 1438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      43581 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGAGT 43522

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43521 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 43462

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43461 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 43402

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43401 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 43342

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43341 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 43282

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43281 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 43222
```

```
Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43221 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 43162

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43161 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 43102

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43101 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 43042

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43041 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 42982

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000
              |||||||||||||||||||||||
Db      42981 TTTCTGCAAAATAGATGTTGTG 42960
```

```
RESULT 5
AX033912
LOCUS       AX033912               41008 bp    DNA     linear   PAT 21-SEP-2000
DEFINITION  Sequence 4 from Patent WO9851790.
ACCESSION   AX033912
VERSION     AX033912.1  GI:10280480
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Cancilla,M.R., Choo,K.H. and Du,S.D.
  TITLE     A novel nucleic acid molecule
  JOURNAL   Patent: WO 9851790-A 4 19-NOV-1998;
            CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
            DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES             Location/Qualifiers
     source          1. .41008
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             80.6%;  Score 1611.4;  DB 6;  Length 41008;
  Best Local Similarity   97.4%;  Pred. No. 2e-240;
  Matches 1666;  Conservative    2;  Mismatches   40;  Indels    3;  Gaps    3;

Qy        292 GATTGATCTGTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAAT 351
              | ||||| |||   || | | || || | ||| |||| ||  | |||| |||||  |||
Db          2 GGTTGATTTGTNNATAAGGAAGTTTGGAATCAATCCCGGAAGGAATTTTTTTTTTAAAAA 61

Qy        352 ATTTTTCTGAATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTT 411
              ||||||  ||| | || || ||:||| | |:||| | ||| |||||||||| ||| | |||
Db         62 ATTTTTTGGAAGG-GTTTGGTAWTAAAAAARCCAATTTGGGTTTTTAAAAATAGGAATTT 120

Qy        412 TATGGTATATATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATAC 471
              ||||| | | | |||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 TATGGGAAAAAATTTTCCCTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATAC 180

Qy        472 TTAAAGTGGGTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGAT 531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTAAAGTGGGTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGAT 240

Qy        532 AATCTCAACCTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATG 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AATCTCAACCTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATG 300

Qy        592 GTTTGATTTGCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTTCATTGTTC 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTTTGATTTGCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTTCATTGTTC 360
```

```
Qy       652 TTTTTTCATCTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATC 711
Db       361 TTTTTTCATCTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATC 420

Qy       712 TATTGGCTTTTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAA 771
Db       421 TATTGGCTTTTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAA 480

Qy       772 TATACATCTTTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAA 831
Db       481 TATACATCTTTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAA 540

Qy       832 TGTAGAAACCTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTA-TGTCT 890
Db       541 TGTAGAAACCTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATTGTCT 600

Qy       891 ATAATACATTAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCA 950
Db       601 ATAATACATTAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCA 660

Qy       951 AACATTTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGT 1010
Db       661 AACATTTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGT 720

Qy      1011 TCCATACCACCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCAC 1070
Db       721 TCCATACCACCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCAC 780

Qy      1071 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 1130
Db       781 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 840

Qy      1131 TAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATT 1190
Db       841 TCAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATT 900

Qy      1191 ACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGA 1250
Db       901 ACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGA 960

Qy      1251 AGGTCTTTTCTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGG 1310
Db       961 AGGTCTTTTCTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGG 1020

Qy      1311 CAGTTTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGA 1370
Db      1021 CAGTTTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGA 1080

Qy      1371 TTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAA 1430
Db      1081 TTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAA 1140

Qy      1431 AATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATT 1489
Db      1141 AATTGGAGTCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATT 1200

Qy      1490 CTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCA 1549
Db      1201 CTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCA 1260

Qy      1550 TCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAA 1609
Db      1261 TCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAA 1320

Qy      1610 GTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATT 1669
Db      1321 GTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATT 1380

Qy      1670 CCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACC 1729
Db      1381 CCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACC 1440

Qy      1730 TCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATA 1789
Db      1441 TCTCCAAGTCATCCATGAGGGCTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATA 1500

Qy      1790 TTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCC 1849
Db      1501 TTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCC 1560

Qy      1850 AAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCC 1909
Db      1561 AAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCC 1620

Qy      1910 AGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCC 1969
Db      1621 AGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCC 1680

Qy      1970 TTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
Db      1681 TTGATCCATTTTCTGCAAAATAGATGTTGTG 1711
```

```
RESULT 6
AF222856
LOCUS       AF222856               69058 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone PnC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222856
VERSION     AF222856.1  GI:9246846
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 69058)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 69058)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Human genomic sequence from 10q25.2, clonelib=PnC
FEATURES             Location/Qualifiers
     source          1. .69058
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="PnC"
     repeat_region   complement(598. .3183)
                     /rpt_family="Tigger1"
     repeat_region   2587. .2715
                     /rpt_family="Alu"
     repeat_region   3283. .3553
                     /rpt_family="Alu"
     repeat_region   complement(4548. .4809)
                     /rpt_family="Alu"
     repeat_region   complement(5919. .6211)
                     /rpt_family="Alu"
     repeat_region   6257. .6599
                     /rpt_family="MLT2B2"
     repeat_region   complement(8848. .9123)
                     /rpt_family="Alu"
     repeat_region   complement(10913. .11198)
                     /rpt_family="Alu"
     repeat_region   complement(12062. .12313)
                     /rpt_family="Alu"
```

```
     repeat_region   complement(13957. .14236)
                     /rpt_family="Alu"
     repeat_region   17133. .17424
                     /rpt_family="Alu"
     repeat_region   complement(19882. .20197)
                     /rpt_family="Alu"
     repeat_region   complement(22664. .22956)
                     /rpt_family="Alu"
     repeat_region   complement(23935. .24113)
                     /rpt_family="MIR"
     repeat_region   24585. .24884
                     /rpt_family="Alu"
     repeat_region   complement(25618. .25973)
                     /rpt_family="THE1"
     repeat_region   26732. .26970
                     /rpt_family="MER2"
     repeat_region   27173. .27449
                     /rpt_family="Alu"
     repeat_region   27526. .27796
                     /rpt_family="Alu"
     repeat_region   28702. .28978
                     /rpt_family="Alu"
     repeat_region   complement(30668. .30736)
                     /rpt_family="MLT1"
     repeat_region   complement(31460. .31993)
                     /rpt_family="MER1"
     repeat_region   complement(32130. .32248)
                     /rpt_family="MLT1"
     repeat_region   32685. .32961
                     /rpt_family="Alu"
     repeat_region   33113. .33403
                     /rpt_family="Alu"
     repeat_region   complement(34169. .34233)
                     /rpt_family="MIR"
     repeat_region   34332. .34639
                     /rpt_family="Alu"
     repeat_region   complement(35993. .36081)
                     /rpt_family="MIR"
     repeat_region   complement(38438. .38721)
                     /rpt_family="Alu"
     repeat_region   39081. .39333
                     /rpt_family="Alu"
     repeat_region   complement(42817. .42877)
                     /rpt_family="MIR"
     repeat_region   complement(44845. .45119)
                     /rpt_family="Alu"
     repeat_region   45532. .45665
                     /rpt_family="MIR"
     repeat_region   complement(46257. .46490)
                     /rpt_family="Alu"
     repeat_region   48262. .48578
                     /rpt_family="Alu"
     repeat_region   complement(48687. .50337)
                     /rpt_family="L1"
     repeat_region   complement(50349. .50611)
                     /rpt_family="Alu"
     repeat_region   complement(50670. .50831)
                     /rpt_family="L1"
     repeat_region   complement(51085. .51374)
                     /rpt_family="Alu"
     repeat_region   complement(52905. .53197)
                     /rpt_family="Alu"
     repeat_region   53896. .54467
                     /rpt_family="LTR12"
     repeat_region   59129. .59723
                     /rpt_family="LTR12"
     repeat_region   60018. .60302
                     /rpt_family="Alu"
     repeat_region   complement(60338. .60470)
                     /rpt_family="MIR"
     repeat_region   complement(61353. .61521)
                     /rpt_family="MER5"
     repeat_region   63135. .63416
                     /rpt_family="Alu"
     repeat_region   63724. .63889
                     /rpt_family="MER5"
     repeat_region   64107. .64182
                     /rpt_family="MIR"
     repeat_region   65459. .65548
                     /rpt_family="MER5"
     repeat_region   66439. .66618
                     /rpt_family="MER20"
     repeat_region   complement(67562. .68084)
                     /rpt_family="MER41"
ORIGIN

  Query Match            79.6%;  Score 1592.8;  DB 8;  Length 69058;
  Best Local Similarity  99.8%;  Pred. No. 1.2e-237;
  Matches 1616;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy       383 CCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATATATTTTTCCTTTTTTTTTTTTT 442
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 CCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATATATTTTTCCTTTTTTTTTTTTT 60

Qy       443 TAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGGGTGTCTTATAGGCAGCATATAT 502
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 TAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGGGTGTCTTATAGGCAGCATATAT 120

Qy       503 CTGGGTCTTGATGTATTATTTAATCTGATAATCTCAACCTTTTTGTTGGAGTGTTTAGGC 562
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CTGGGTCTTGATGTATTATTTAATCTGATAATCTCAACCTTTTTGTTGGAGTGTTTAGGC 180

Qy       563 CATTTACATTTAGTGTAATTATAGACATGGTTTGATTTGCTATACCATCTTTTCATTTGT 622
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 CATTTACATTTAGTGTAATTATAGACATGGTTTGATTTGCTATACCATCTTTTCATTTGT 240

Qy       623 TTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCATCTTTGACCATTTTCTTTAGTAC 682
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCATCTTTGACCATTTTCTTTAGTAC 300

Qy       683 TGAATACTTTTTTTGTATTTCATTATATCTATTGGCTTTTTAGTTATACCTCTTAAAATT 742
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TGAATACTTTTTTTGTATTTCATTATATCTATTGGCTTTTTAGTTATACCTCTTAAAATT 360

Qy       743 TTTTTTTCTGTTTTATGTAGGATTTATAATATACATCTTTAACTTATCACAGATTACCTT 802
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TTTTTTTCTGTTTTATGTAGGATTTATAATATACATCTTTAACTTATCACAGATTACCTT 420

Qy       803 CAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAACCTTACAAGAGTATATTTTCATT 862
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAACCTTACAAGAGTATATTTTCATT 480

Qy       863 TCTGTCTCCTAATTTTTATGCTA-TGTCTATAATACATTAGGTTTGTTGTTGTTTGTTTT 921
            |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       481 TCTGTCTCCTAATTTTTATGCTATTGTCTATAATACATTAGGTTTGTTGTTGTTTGTTTT 540

Qy       922 TACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCTGTAAAGGGCTAGATAGTACAG 981
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCTGTAAAGGGCTAGATAGTACAG 600

Qy       982 GCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAA 1041
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 GCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAA 660

Qy      1042 GAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAA 1101
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAA 720

Qy      1102 AGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAA 1161
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       721 AGTTTTGCTTATACTACACTGTAGTCTGTTCAGTGTGCAATAGTGTTATGTCTAAAAAAA 780

Qy      1162 CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATT 1221
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATT 840
```

```
Qy       1222 CAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGG 1281
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGG 900

Qy       1282 TCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAA 1341
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 TCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAA 960

Qy       1342 TATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAG 1401
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 TATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAG 1020

Qy       1402 CATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCT 1460
             ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1021 CATTTTATGCACAGTAGAACTTCTTTGAAAATTGGAGTCAATCCTCTCAAACCCTGCTCT 1080

Qy       1461 GCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTT 1520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTT 1140

Qy       1521 CACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCAT 1580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCAT 1200

Qy       1581 CCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCA 1640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCA 1260

Qy       1641 TGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGT 1700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGT 1320

Qy       1701 GGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGA 1760
             ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db       1321 GGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGCTGGAATCGA 1380

Qy       1761 CTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTTCTT 1820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTTCTT 1440

Qy       1821 AATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCC 1880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCC 1500

Qy       1881 ATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAA 1940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 ATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAA 1560

Qy       1941 TAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 1620
```

```
RESULT 7
AF222854
LOCUS       AF222854               80202 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222854
VERSION     AF222854.1  GI:9246844
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80202)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 80202)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Human genomic sequence from 10q25.2, clonelib=NC, second release
FEATURES             Location/Qualifiers
     source          1. .80202
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="NC"
     repeat_region   complement(598. .3183)
                     /rpt_family="Tigger1"
     repeat_region   2587. .2715
                     /rpt_family="Alu"
     repeat_region   3283. .3553
                     /rpt_family="Alu"
     repeat_region   complement(4548. .4809)
                     /rpt_family="Alu"
     repeat_region   complement(5919. .6211)
                     /rpt_family="Alu"
     repeat_region   6257. .6599
                     /rpt_family="MLT2B2"
     repeat_region   complement(8848. .9123)
                     /rpt_family="Alu"
     repeat_region   complement(10913. .11198)
                     /rpt_family="Alu"
     repeat_region   complement(12062. .12313)
                     /rpt_family="Alu"
     repeat_region   complement(13957. .14236)
                     /rpt_family="Alu"
     repeat_region   17133. .17424
                     /rpt_family="Alu"
     repeat_region   complement(19882. .20197)
                     /rpt_family="Alu"
     repeat_region   complement(22664. .22956)
                     /rpt_family="Alu"
     repeat_region   complement(23935. .24113)
                     /rpt_family="MIR"
     repeat_region   24585. .24884
                     /rpt_family="Alu"
     repeat_region   complement(25618. .25973)
                     /rpt_family="THE1"
     repeat_region   26732. .26970
                     /rpt_family="MER2"
     repeat_region   27173. .27449
                     /rpt_family="Alu"
     repeat_region   27526. .27796
                     /rpt_family="Alu"
     repeat_region   28702. .28978
                     /rpt_family="Alu"
     repeat_region   complement(30668. .30736)
                     /rpt_family="MLT1"
     repeat_region   complement(31460. .31993)
                     /rpt_family="MER1"
     repeat_region   complement(32130. .32248)
                     /rpt_family="MLT1"
     repeat_region   32685. .32961
                     /rpt_family="Alu"
     repeat_region   33113. .33403
                     /rpt_family="Alu"
     repeat_region   complement(34169. .34233)
                     /rpt_family="MIR"
     repeat_region   34332. .34639
                     /rpt_family="Alu"
     repeat_region   complement(35993. .36081)
                     /rpt_family="MIR"
     repeat_region   complement(38438. .38721)
                     /rpt_family="Alu"
```

```
     repeat_region   39081. .39333
                     /rpt_family="Alu"
     repeat_region   complement(42817. .42877)
                     /rpt_family="MIR"
     repeat_region   complement(44845. .45119)
                     /rpt_family="Alu"
     repeat_region   45532. .45665
                     /rpt_family="MIR"
     repeat_region   complement(46257. .46490)
                     /rpt_family="Alu"
     repeat_region   48262. .48578
                     /rpt_family="Alu"
     repeat_region   complement(48687. .50337)
                     /rpt_family="L1"
     repeat_region   complement(50349. .50611)
                     /rpt_family="Alu"
     repeat_region   complement(50670. .50831)
                     /rpt_family="L1"
     repeat_region   complement(51085. .51374)
                     /rpt_family="Alu"
     repeat_region   complement(52905. .53197)
                     /rpt_family="Alu"
     repeat_region   53896. .54467
                     /rpt_family="LTR12"
     repeat_region   59129. .59719
                     /rpt_family="LTR12"
     repeat_region   60018. .60302
                     /rpt_family="Alu"
     repeat_region   complement(60338. .60470)
                     /rpt_family="MIR"
     repeat_region   complement(61353. .61521)
                     /rpt_family="MER5"
     repeat_region   63135. .63416
                     /rpt_family="Alu"
     repeat_region   63724. .63889
                     /rpt_family="MER5"
     repeat_region   64107. .64182
                     /rpt_family="MIR"
     repeat_region   65459. .65548
                     /rpt_family="MER5"
     repeat_region   66439. .66618
                     /rpt_family="MER20"
     repeat_region   complement(67562. .68084)
                     /rpt_family="MER41"
     repeat_region   complement(72861. .73139)
                     /rpt_family="Alu"
     repeat_region   73756. .74037
                     /rpt_family="Alu"
     repeat_region   75588. .75853
                     /rpt_family="Alu"
     repeat_region   complement(77441. .77729)
                     /rpt_family="Alu"
ORIGIN

  Query Match             79.6%;  Score 1592.8;  DB 8;  Length 80202;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-237;
  Matches 1616;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy        383 CCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATATATTTTTCCTTTTTTTTTTTTT 442
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATATATTTTTCCTTTTTTTTTTTTT 60

Qy        443 TAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGGGTGTCTTATAGGCAGCATATAT 502
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGGGTGTCTTATAGGCAGCATATAT 120

Qy        503 CTGGGTCTTGATGTATTATTTAATCTGATAATCTCAACCTTTTTGTTGGAGTGTTTAGGC 562
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGGTCTTGATGTATTATTTAATCTGATAATCTCAACCTTTTTGTTGGAGTGTTTAGGC 180

Qy        563 CATTTACATTTAGTGTAATTATAGACATGGTTTGATTTGCTATACCATCTTTTCATTTGT 622
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATTTACATTTAGTGTAATTATAGACATGGTTTGATTTGCTATACCATCTTTTCATTTGT 240

Qy        623 TTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCATCTTTGACCATTTTCTTTAGTAC 682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCATCTTTGACCATTTTCTTTAGTAC 300

Qy        683 TGAATACTTTTTTTGTATTTCATTATATCTATTGGCTTTTTAGTTATACCTCTTAAAATT 742
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAATACTTTTTTTGTATTTCATTATATCTATTGGCTTTTTAGTTATACCTCTTAAAATT 360

Qy        743 TTTTTTTCTGTTTTATGTAGGATTTATAATATACATCTTTAACTTATCACAGATTACCTT 802
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTTTTTTCTGTTTTATGTAGGATTTATAATATACATCTTTAACTTATCACAGATTACCTT 420

Qy        803 CAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAACCTTACAAGAGTATATTTTCATT 862
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAACCTTACAAGAGTATATTTTCATT 480

Qy        863 TCTGTCTCCTAATTTTTATGCTA-TGTCTATAATACATTAGGTTTGTTGTTGTTTGTTTT 921
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        481 TCTGTCTCCTAATTTTTATGCTATTGTCTATAATACATTAGGTTTGTTGTTGTTTGTTTT 540

Qy        922 TACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCTGTAAAGGGCTAGATAGTACAG 981
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCTGTAAAGGGCTAGATAGTACAG 600

Qy        982 GCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAA 1041
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAA 660

Qy       1042 GAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAA 1101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAA 720

Qy       1102 AGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAA 1161
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        721 AGTTTTGCTTATACTACACTGTAGTCTGTTCAGTGTGCAATAGTGTTATGTCTAAAAAAA 780

Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATT 1221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATT 840

Qy       1222 CAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGG 1281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGG 900

Qy       1282 TCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAA 1341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAA 960

Qy       1342 TATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAG 1401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAG 1020

Qy       1402 CATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCT 1460
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db       1021 CATTTTATGCACAGTAGAACTTCTTTGAAAATTGGAGTCAATCCTCTCAAACCCTGCTCT 1080

Qy       1461 GCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTT 1520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTT 1140

Qy       1521 CACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCAT 1580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCAT 1200

Qy       1581 CCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCA 1640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCA 1260

Qy       1641 TGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGT 1700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGT 1320
```

```
Qy       1701 GGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGA 1760
             ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1321 GGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGCTGGAATCGA 1380

Qy       1761 CTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTTCTT 1820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTTCTT 1440

Qy       1821 AATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCC 1880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCC 1500

Qy       1881 ATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAA 1940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 ATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAA 1560

Qy       1941 TAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 1620


RESULT 8
AF042484
LOCUS       AF042484               80155 bp    DNA     linear   PRI 22-JUN-2000
DEFINITION  Homo sapiens sequence of the 80 kb core centromere protein binding
            region of the mar del(10) neocentromere corresponding to human
            chromosome 10q25.2, complete sequence.
ACCESSION   AF042484
VERSION     AF042484.1  GI:4205782
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,K.H.
  TITLE     Sequence analysis of an 80 kb human neocentromere
  JOURNAL   Hum. Mol. Genet. 8 (2), 217-227 (1999)
   PUBMED   9931329
REFERENCE   2  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville, Melbourne, Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .80155
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /note="Sequence information obtained from BAC clones
                     prepared from cell line BE2C1-18-5f"
     repeat_region   complement(join(49. .525,3628. .4545,4849. .5053))
                     /rpt_family="L1MEc"
                     /rpt_type=dispersed
     repeat_region   join(560. .597,3184. .3281)
                     /rpt_family="MER58"
                     /rpt_type=dispersed
     repeat_region   complement(join(598. .2553,2716. .3183))
                     /rpt_family="TIGGER1"
                     /rpt_type=dispersed
     repeat_region   2558. .2581
                     /standard_name="T1"
                     /note="100% homology"
                     /rpt_type=direct
                     /rpt_unit="(caggcacagtgg)2"
     repeat_region   2570. .2715
                     /rpt_family="AluSg/x"
                     /rpt_type=dispersed
     repeat_region   3021. .3050
                     /standard_name="T2"
                     /note="83.3% homology"
                     /rpt_type=direct
                     /rpt_unit="(taacaaagtg)3"
     repeat_region   3283. .3558
                     /rpt_family="AluSx"
                     /rpt_type=dispersed
     repeat_region   4173. .4211
                     /standard_name="T3"
                     /note="82.1% homology"
                     /rpt_type=direct
                     /rpt_unit="(aaaaaaataattt)3"
     repeat_region   complement(4546. .4848)
                     /rpt_family="AluSx"
                     /rpt_type=dispersed
     repeat_region   complement(5909. .6208)
                     /rpt_family="AluSg"
                     /rpt_type=dispersed
     repeat_region   join(6232. .6573,6638. .6688)
                     /rpt_family="MLT2CB"
                     /rpt_type=dispersed
     repeat_region   6598. .6645
                     /standard_name="AC24"
                     /note="100% homology"
                     /rpt_type=tandem
                     /rpt_unit="ac"
     repeat_region   6598. .6637
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ca"
     repeat_region   7091. .7134
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   7808. .7911
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   8349. .8524
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   complement(8839. .9124)
                     /rpt_family="AluJb"
                     /rpt_type=dispersed
     repeat_region   9180. .9280
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   10070. .10170
                     /note="low complexity"
                     /rpt_family="purine-rich"
     repeat_region   10097. .10136
                     /standard_name="GA1"
                     /note="82.5% homology"
                     /rpt_type=direct
                     /rpt_unit="(gaaagaaagg)4"
     repeat_region   10175. .10339
                     /rpt_family="FRAM"
                     /rpt_type=dispersed
     repeat_region   complement(10899. .11197)
                     /rpt_family="AluSx"
                     /rpt_type=dispersed
     misc_feature    join(11627. .11757,12434. .12512,13443. .13562)
                     /note=">98% identity"
                     /db_xref="dbEST:AA774571"
     repeat_region   11996. .12028
                     /standard_name="AT4"
                     /note="100% homology"
                     /rpt_type=tandem
                     /rpt_unit="(attt)8"
     repeat_region   complement(12010. .12312)
                     /rpt_family="AluJb"
                     /rpt_type=dispersed
     repeat_region   12327. .12433
                     /rpt_family="L2"
                     /rpt_type=dispersed
```

```
repeat_region   complement(13946. .14235)
                /rpt_family="AluSq"
                /rpt_type=dispersed
repeat_region   15130. .15159
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   15164. .15753
                /standard_name="AT28"
                /note="82.1% homology"
                /rpt_type=direct
                /rpt_unit="(atgtatatatgtgtatatagacataaat)21"
repeat_region   join(15170. .15289,15410. .15612)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="ta"
repeat_region   15617. .15731
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="cata"
repeat_region   16004. .16100
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16128. .16364)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16407. .16592)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16657. .16889)
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   16914. .16936
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   17102. .17386
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   17406. .17564
                /standard_name="GA2"
                /note="81.8% homology"
                /rpt_type=direct
repeat_region   17416. .17543
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   17647. .17702
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   complement(19863. .20161)
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   complement(20842. .20913)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(21366. .21414)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   21375. .21639
                /standard_name="CT1"
                /note="75% homology"
                /rpt_type=dispersed
repeat_region   complement(21415. .21536)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="tggaa"
repeat_region   complement(21524. .21646)
                /note="low complexity"
                /rpt_family="purine-rich"
repeat_region   22409. .22646
                /rpt_family="L1M4"
                /rpt_type=dispersed
repeat_region   22560. .22604
                /standard_name="T4"
                /note="80% homology"
                /rpt_type=direct
                /rpt_unit="(aatattacaataatt)3"

  Query Match            78.8%;  Score 1575.2;  DB 8;  Length 80155;
  Best Local Similarity  99.1%;  Pred. No. 6.1e-235;
  Matches 1605;  Conservative    0;  Mismatches   13;  Indels    2;  Gaps    2;

Qy       383 CCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATATATTTTTCCTTTTTTTTTTTTTT 442
            ||| | ||| ||||||||| ||| | |||||||| | | | |||||| ||||||||||||
Db         1 CCAATTTGGGTTTTTAAAAATAGGAATTTTATGGGAAAAAATTTTCCCTTTTTTTTTTTTT 60

Qy       443 TAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGGGTGTCTTATAGGCAGCATATAT 502
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 TAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGGGTGTCTTATAGGCAGCATATAT 120

Qy       503 CTGGGTCTTGATGTATTATTTAATCTGATAATCTCAACCTTTTTGTTGGAGTGTTTAGGC 562
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CTGGGTCTTGATGTATTATTTAATCTGATAATCTCAACCTTTTTGTTGGAGTGTTTAGGC 180

Qy       563 CATTTACATTTAGTGTAATTATAGACATGGTTTGATTTGCTATACCATCTTTTCATTTGT 622
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 CATTTACATTTAGTGTAATTATAGACATGGTTTGATTTGCTATACCATCTTTTCATTTGT 240

Qy       623 TTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCATCTTTGACCATTTTCTTTAGTAC 682
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCATCTTTGACCATTTTCTTTAGTAC 300

Qy       683 TGAATACTTTTTTTGTATTTCATTATATCTATTGGCTTTTTAGTTATACCTCTTAAAATT 742
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TGAATACTTTTTTTGTATTTCATTATATCTATTGGCTTTTTAGTTATACCTCTTAAAATT 360

Qy       743 TTTTTTTCTGTTTTATGTAGGATTTATAATATACATCTTTAACTTATCACAGATTACCTT 802
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TTTTTTTCTGTTTTATGTAGGATTTATAATATACATCTTTAACTTATCACAGATTACCTT 420

Qy       803 CAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAACCTTACAAGAGTATATTTTCATT 862
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAACCTTACAAGAGTATATTTTCATT 480

Qy       863 TCTGTCTCCTAATTTTTATGCTA-TGTCTATAATACATTAGGTTTGTTGTTGTTTGTTTT 921
            |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       481 TCTGTCTCCTAATTTTTATGCTATTGTCTATAATACATTAGGTTTGTTGTTGTTTGTTTT 540

Qy       922 TACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCTGTAAAGGGCTAGATAGTACAG 981
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCTGTAAAGGGCTAGATAGTACAG 600

Qy       982 GCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAA 1041
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 GCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAA 660

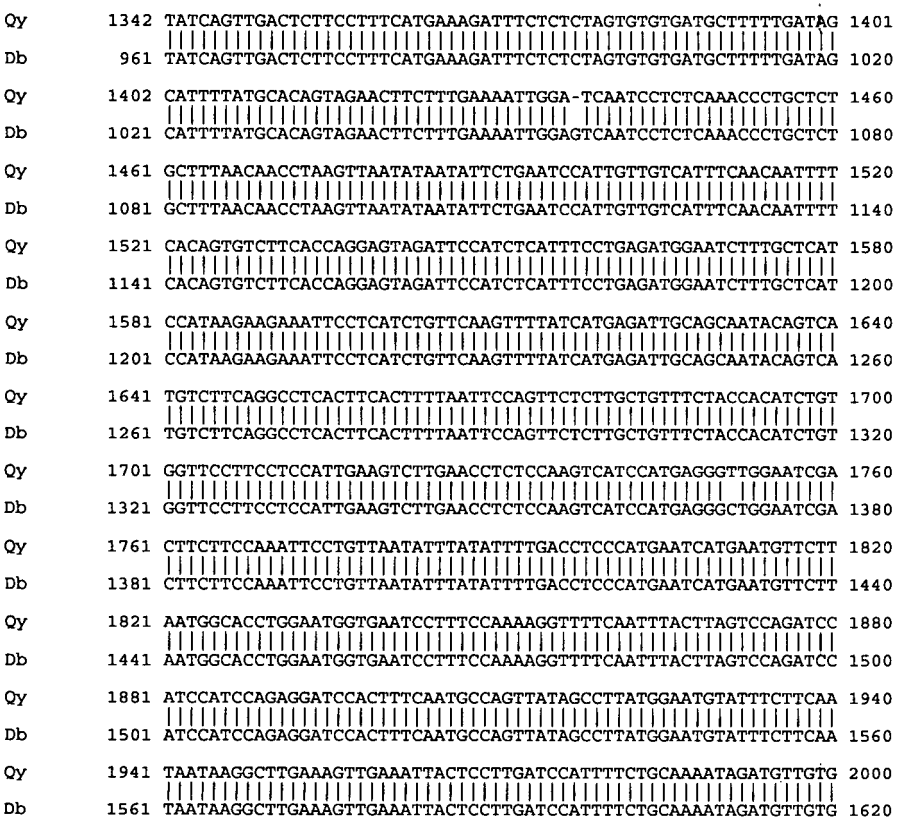
Qy      1042 GAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAA 1101
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAA 720

Qy      1102 AGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAA 1161
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       721 AGTTTTGCTTATACTACACTGTAGTCTGTTCAGTGTGCAATAGTGTTATGTCTAAAAAAA 780

Qy      1162 CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATT 1221
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATT 840

Qy      1222 CAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGG 1281
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 CAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGG 900

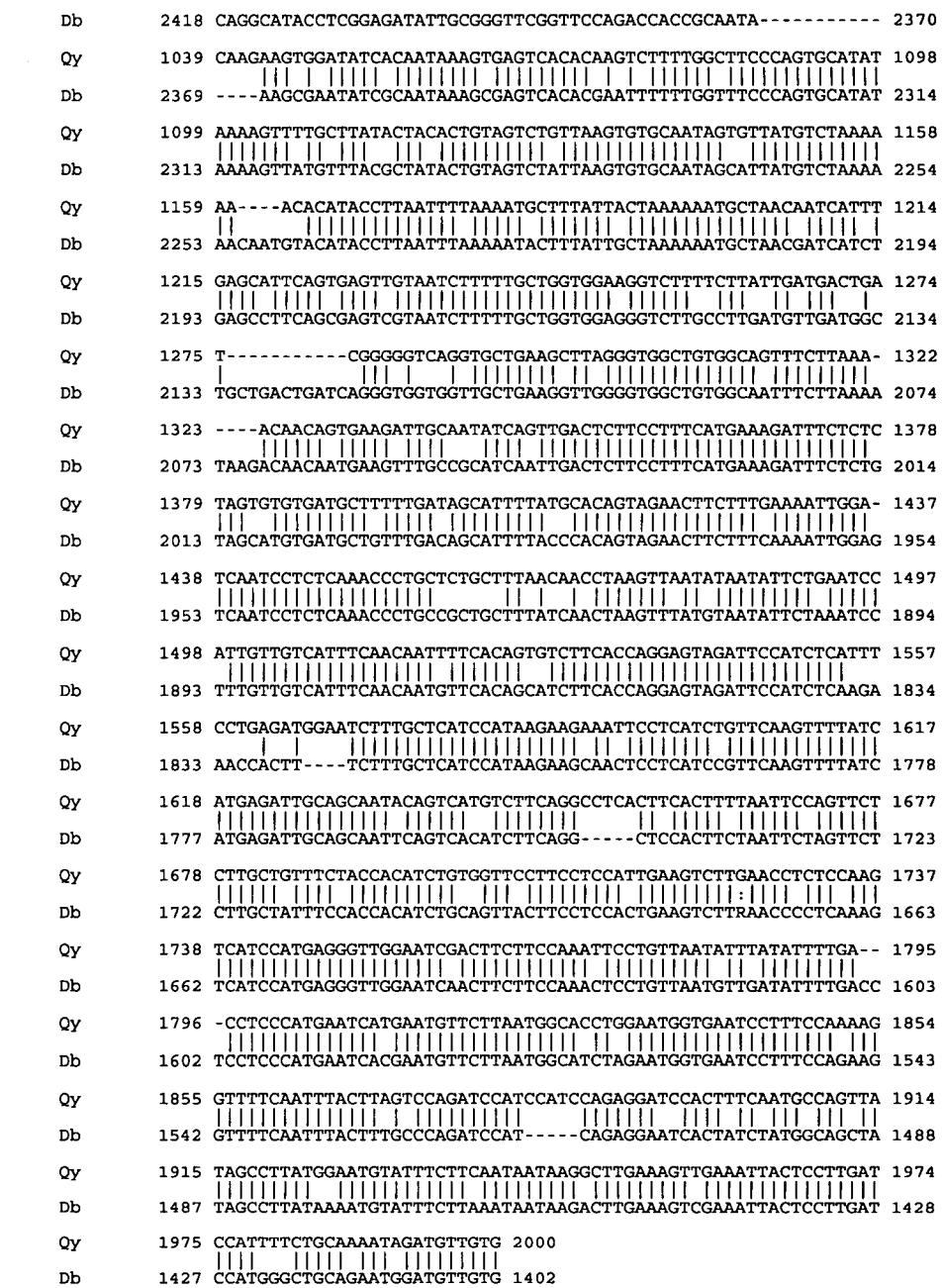
Qy      1282 TCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAA 1341
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 TCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAA 960
```



```
RESULT 9
AR612649/c
LOCUS       AR612649               2418 bp    DNA     linear   PAT 15-DEC-2004
DEFINITION  Sequence 148 from patent US 6828097.
ACCESSION   AR612649
VERSION     AR612649.1  GI:56668471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2418)
  AUTHORS   Knoll,J.H.M. and Rogan,P.K.
  TITLE     Single copy genomic hybridization probes and method of generating
            same
  JOURNAL   Patent: US 6828097-A 148 07-DEC-2004;
            The Childrens Mercy Hospital; Kansas City, MO
FEATURES             Location/Qualifiers
     source          1. .2418
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             31.8%;  Score 636.8;  DB 6;  Length 2418;
  Best Local Similarity   82.1%;  Pred. No. 2.6e-89;
  Matches  859;  Conservative    1;  Mismatches  133;  Indels   53;  Gaps    9;
```





```
RESULT 10
HSU49973/c
```

```
LOCUS       HSU49973                2418 bp    DNA     linear   PRI 08-JUN-2005
DEFINITION  Human Tigger1 transposable element, complete consensus sequence.
ACCESSION   U49973
VERSION     U49973.1  GI:2226003
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2418)
  AUTHORS   Smit,A.F. and Riggs,A.D.
  TITLE     Tiggers and DNA transposon fossils in the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (4), 1443-1448 (1996)
   PUBMED   8643651
REFERENCE   2  (bases 1 to 2418)
  AUTHORS   Robertson,H.M.
  TITLE     Members of the pogo superfamily of DNA-mediated transposons in the
            human genome
  JOURNAL   Mol. Gen. Genet. 252 (6), 761-766 (1996)
   PUBMED   8917322
REFERENCE   3  (bases 1 to 2418)
  AUTHORS   Robertson,H.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-1996) Hugh M. Robertson, Entomology, University
            of Illinois at Urbana-Champaign, 505 S. Goodwin, Urbana, IL 61801,
            USA
COMMENT     On Jun 28, 1997 this sequence version replaced gi:1224064.
FEATURES             Location/Qualifiers
     source          1. .2418
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     repeat_region   1. .2418
                     /note="consensus sequence based on 50 full-length genomic
                     sequences"
                     /transposon="Tigger1"
     repeat_region   1. .13
                     /rpt_type=inverted
     CDS             425. .1789
                     /note="ORF1; MER37; similar to pogo element"
                     /citation=[1]
                     /codon_start=1
                     /product="putative transposase"
                     /protein_id="AAB61714.1"
                     /db_xref="GI:2226004"
                     /translation="MASKCSSERKSRTSLTLNQKLEMIKLSEEGMSKAEIGRKLGLLR
                     QTVSQVVNAKEKFLKEIKSATPVNTRMIRKRNSLIADMEKVLVVWIEDQTSHNIPLSQ
                     SLIQSKALTLFNSMKAERGEEAAEEKLEASRGWFMRFKERSRLHNIKVQGEAASADGE
                     AAASYPEDLAKIIDEGGYTKQQIFNVDETAFYWKKMPSRTFIAREEKSMPGFKASKDR
                     LTLLLGANAAGDFKLKPMLIYHSENPRALKNYAKSTLPVLYKWNNKAWMTAHLFTAWF
                     TEYFKPTVETYCSEKKISFKILLLIDNAPGHPRALMEMYKEINVVFMPANTTSILQPM
                     DQGVISTFKSYYLRNTFRKAIAAIDSDSSDGSGQSKLKTFWKGFTILDAIKNIRDSWE
                     EVKISTLTGVWKKLIPTLMDDFEGFKTSVEEVTADVVEIARELELEVEPEDVTELLQS
                     HDKT"
     CDS             1811. .2206
                     /note="ORF2: function unknown"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAB61715.1"
                     /db_xref="GI:2226005"
                     /translation="MDEQRKWFLEMESTPGEDAVNIVEMTTKDLEYYINLVDKAAAGF
                     ERIDSNFERSSTVGKMLSNSIACYREIFRERKSQSMRQTSLLSYFKKLPQPPQPSATT
                     TLISQQPSTSRQDPPPAKRLRLAEGSDDR"
     polyA_signal    2218. .2223
     repeat_region   2405. .2418
                     /rpt_type=inverted
ORIGIN

  Query Match           31.7%;  Score 634;  DB 8;  Length 2418;
  Best Local Similarity   82.0%;  Pred. No. 7.1e-89;
  Matches  858;  Conservative    0;  Mismatches  135;  Indels   53;  Gaps    9;
```

```
Qy        979 CAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATG 1038
              |||||||||| |||||| || ||||| ||||||| |||||| ||||||  |||
Db       2418 CAGGCATACCTCGGAGATATTGCGGGTTCGGTTCCAGACCACCGCAATAAAGCAA----- 2364

Qy       1039 CAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATAT 1098
                  ||||| |||||||| ||||||||| | | |||||| ||||||||||||||
Db       2363 ----------ATATCGCAATAAAGCGAGTCACACGAATTTTTGGTTTCCCAGTGCATAT 2314

Qy       1099 AAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAA 1158
              ||||||| || ||| |||| |||||||||| ||||||||||||||  ||||||||||
Db       2313 AAAAGTTATGTTTACACTATACTGTAGTCTATTAAGTGTGCAATAGCATTATGTCTAAAA 2254

Qy       1159 AA----ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTT 1214
              ||     |||||||||||||| ||||| ||||||| |||||||||||||| ||||| |
Db       2253 AACAATGTACATACCTTAATTTAAAAATACTTTATTGCTAAAAAATGCTAACGATCATCT 2194

Qy       1215 GAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGA 1274
              |||| ||||| |||| |||||||||||||||||||| ||||||  ||   || |||  |
Db       2193 GAGCCTTCAGCGAGTCGTAATCTTTTTGCTGGTGGAGGGTCTTGCCTCGATGTTGATGGC 2134

Qy       1275 T-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA- 1322
              |           ||| |   | |||||||| || ||||||||||||| |||||||||
Db       2133 TGCTGACTGATCAGGGTGGTGGTTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCTTAAAA 2074

Qy       1323 ----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTC 1378
                  |||||| ||||| ||||   ||| |||||||||||||||| ||||||||||||
Db       2073 TAAGACAACAATGAAGTTTGCCGCATCGATTGACTCTTCCTTTCACGAAAGATTTCTCTG 2014

Qy       1379 TAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA- 1437
              |||  || |||||| |||||||||||||||  ||||||||||||||||| |||||||||
Db       2013 TAGCATGCGATGCTGTTTGATAGCATTTTACCCACAGTAGAACTTCTTTCAAAATTGGAG 1954

Qy       1438 TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCC 1497
              |||||||||||||||||||||      || |  | ||||||| || |||||||||| |||||
Db       1953 TCAATCCTCTCAAACCCTGCCGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCC 1894

Qy       1498 ATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTT 1557
              |||||||||||||||||| |||||||  |||||||||||||||||||||||||||||
Db       1893 TTTGTTGTCATTTCAACAATGTTCACAGCATCTTCACCAGGAGTAGATTCCATCTCAAGA 1834

Qy       1558 CCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATC 1617
                |  |    |||||||||||||||||||| || |||||||| ||||||||||||||
Db       1833 AACCACTT----TCTTTGCTCATCCATAAGAAGCAACTCCTCATCCGTTCAAGTTTTATC 1778

Qy       1618 ATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCT 1677
              |||||||||||||||| ||||||  ||||||||      || ||||| ||||||  ||||||
Db       1777 ATGAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCCACTTCTAATTCTAGTTCT 1723

Qy       1678 CTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAG 1737
              |||||| |||| |||||||||  ||| |||||||| |||||||||||||| ||| |||
Db       1722 CTTGCTATTTCCACCACATCTGCAGTTACTTCCTCCACTGAAGTCTTGAACCCCTCAAAG 1663

Qy       1738 TCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA-- 1795
              |||||||||||||||||||| |||||||||| |||||||||  || |||||||||
Db       1662 TCATCCATGAGGGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGATATTTTGACC 1603

Qy       1796 -CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAG 1854
              ||||||||||||||| |||||||||||||||| || |||||||||||||||||| |||
Db       1602 TCCTCCCATGAATCACGAATGTTCTTAATGGCATCTAGAATGGTGAATCCTTTCCAGAAG 1543

Qy       1855 GTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTA 1914
              |||||||||||||| | ||||||||||     |||||||  |||| || ||| ||| ||
Db       1542 GTTTTCAATTTACTTTGCCCAGATCCAT-----CAGAGGAATCACTATCTATGGCAGCTA 1488

Qy       1915 TAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGAT 1974
              |||||||| | |||||||||||| ||||||||| ||||||||| ||||||||||||||||
Db       1487 TAGCCTTACGAAATGTATTTCTTAAATAATAAGACTTGAAAGTCGAAATTACTCCTTGAT 1428

Qy       1975 CCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||   ||||| ||| |||||||||
Db       1427 CCATGGGCTGCAGAATGGATGTTGTG 1402
```

```
RESULT 11
AC092165
LOCUS       AC092165               100020 bp    DNA     linear   PRI 30-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-753F4 from 2, complete sequence.
ACCESSION   AC092165
VERSION     AC092165.4  GI:18056750
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 100020)
  AUTHORS   Haglund,K. and Abbott,S.
  TITLE     The sequence of Homo sapiens BAC clone RP11-753F4
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 100020)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 100020)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 100020)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 100020)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jan 4, 2002 this sequence version replaced gi:17981677.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0753F04
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-762N20, 2000 bp overlap;
            the clone sequenced to the right is RP11-485D7.  Actual start of
            this clone is at base position 98039 of RP11-762N20; actual end is
            at base position 100020 of RP11-753F4.

            Data from AC09338 and AC074188 was used to finish this clone,
            AC092165. Polymorphisms have been identified between AC074188 and
            AC092165.
FEATURES             Location/Qualifiers
     source          1. .100020
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-753F4"
                     /clone_lib="RPCI-11"
     gene            complement(8894. .16888)
                     /gene="ECEL1"
     mRNA            complement(join(8894. .9319,9466. .9542,9785. .9880,
                     10158. .10223,10573. .10697,10849. .10916,11485. .11536,
                     11617. .11675,11918. .12021,12172. .12246,12483. .12581,
                     13068. .13290,13539. .13663,13868. .13960,14048. .14158,
                     14266. .14334,14935. .15821,16786. .16888))
                     /gene="ECEL1"
     CDS             complement(join(9220. .9319,9466. .9542,9785. .9880,
                     10158. .10223,10573. .10697,10849. .10916,11485. .11536,
                     11617. .11675,11918. .12021,12172. .12246,12483. .12581,
                     13068. .13290,13539. .13663,13868. .13960,14048. .14158,
                     14266. .14334,14935. .15720))
                     /gene="ECEL1"
                     /note="Homo sapiens endothelin converting enzyme-like 1
                     (ECEL1), mRNA.; H_NH0753F04.1
                     This gene was based on gi(4758231)"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAY24101.1"
                     /db_xref="GI:62988714"
                     /translation="MEPPYSLTAHYDEFQEVKYVSRCGAGGARGASLPPGFPLGAARS
                     ATGARSGLPRWNRREVCLLSGLVFAAGLCAILAAMLALKYLGPVAAGGGACPEGCPER
                     KAFARAARFLAANLDASIDPCQDFYSFACGGWLRRHAIPDDKLTYGTIAAIGEQNEER
                     LRRLLARPGGGPGGAAQRKVRAFFRSCLDMREIERLGPRPMLEVIEDCGGWDLGGAEE
                     RPGVAARWDLNRLLYKAQGVYSAAALFSLTVSLDDRNSSRYVIRIDQDGLTLPERTLY
                     LAQDEDSEKILAAYRVFMERVLSLLGADAVEQKAQEILQVEQQLANITVSEHDDLRRD
                     VSSMYNKVTLGQLQKITPHLRWKWLLDQIFQEDFSEEEEVVLLATDYMQQVSQLIRST
                     PHRVLHNYLVWRVVVVLSEHLSPPFREALHELAQEMEGSDKPQELARVCLGQAANRHFG
                     MALGALFVHEHFSAASKAKVQQLVEDIKYILGQRLEELDWMDAETRAAARAKLQYMMV
                     MVGYPDFLLKPDAVDKEYEFEVHEKTYFKNILNSIRFSIQLSVKKIRQEVDKSTWLLP
                     PQALNAYYLPNKNQMVFPAGILQPTLYDPDFPQSLNYGGIGTIIGHELTHGYDDWGGQ
                     YDRSGNLLHWWTEASYSRFLRKAECIVRLYDNFTVYNQRVNGKHTLGENIADMGGLKL
                     AYHAYQKWVREHGPEHPLPRLKYTHDQLFFIAFAQNWCIKRRSQSIYLQVLTDKHAPE
                     HYRVLGSVSQFEEFGRAFHCPKDSPMNPAHKCSVW"
     misc_feature    14636. .17331
                     /note="CpG island (%GC=72.6, o/e=0.83, #CpGs=313)"
     misc_feature    31824. .32934
                     /note="CpG island (%GC=68.7, o/e=0.83, #CpGs=95)"
     misc_feature    50972. .53518
                     /note="CpG island (%GC=70.7, o/e=0.66, #CpGs=204)"
     gene            55279. .64562
                     /gene="CHRND"
     mRNA            join(55279. .55334,55596. .55741,56468. .56512,57329. .57438,
                     57568. .57723,57929. .58038,59006. .59206,60419. .60530,
                     60609. .60723,62998. .63202,63291. .63409,64197. .64562)
                     /gene="CHRND"
     CDS             join(55283. .55334,55596. .55741,56468. .56512,57329. .57438,
                     57568. .57723,57929. .58038,59006. .59206,60419. .60530,
                     60609. .60723,62998. .63202,63291. .63409,64197. .64379)
                     /gene="CHRND"
                     /note="Homo sapiens cholinergic receptor, nicotinic, delta
                     polypeptide (CHRND), mRNA.; H_NH0753F04.2
                     This gene was based on gi(4557460)"
```

```
                /codon_start=1
                /product="unknown"
                /protein_id="AAY24102.1"
                /db_xref="GI:62988715"
                /translation="MEGPVLTLGLLAALAVCGSWGLNEEERLIRHLFQEKGYNKELRP
                VAHKEESVDVALALTLSNLISLKEVEETLTTNVWIEHGWTDNRLKWNAEEFGNISVLR
                LPPDMVWLPEIVLENNNDGSFQISYSCNVLVYHYGFVYWLPPAIFRSSCPISVTYFPF
                DWQNCSLKFSSLKYTAKEITLSLKQDAKENRTYPVEWIIIDPEGFTENGEWEIVHRPA
                RVNVDPRAPLDSPSRQDITFYLIIRRKPLFYIINILVPCVLISFMVNLVFYLPADSGE
                KTSVAISVLLAQSVFLLLISKRLPATSMAIPLIGKFLLFGMVLVTMVVVICVIVLNIH
                FRTPSTHVLSEGVKKLFLETLPELLHMSRPAEDGPSPGALVRRSSSLGYISKAEEYFL
                LKSRSDLMFEKQSERHGLARRLTTARRPPASSEQAQQELFNELKPAVDGANFIVNHMR
                DQNNYNEEKDSWNRVARTVDRLCLFVVTPVMVVGTAWIFLQGVYNQPPPQPFPGDPYS
                YNVQDKRFI"
    gene        68815. .74783
                /gene="CHRNG"
    mRNA        join(68815. .68889,69070. .69198,69448. .69492,69669. .69778,
                70441. .70596,71490. .71587,71949. .72149,72342. .72456,
                72652. .72766,73434. .73647,73839. .73969,74610. .74783)
                /gene="CHRNG"
    CDS         join(68815. .68889,69070. .69198,69448. .69492,69669. .69778,
                70441. .70596,71490. .71587,71949. .72149,72342. .72456,
                72652. .72766,73434. .73647,73839. .73969,74610. .74783)
                /gene="CHRNG"
                /note="Homo sapiens cholinergic receptor, nicotinic, gamma
                polypeptide (CHRNG), mRNA.; H_NH0753F04.3
                This gene was based on gi(7382453)"
                /codon_start=1
                /product="unknown"
                /protein_id="AAY24103.1"
                /db_xref="GI:62988716"
                /translation="MHGGQGPLLLLLLLAVCLGGTQRNLRNQEERLLADLMQNYDPNL
                RPAERDSDVVNVSLKLTLTNLISLNEREEALTTNVWIEMQWCDYRLRWDPRDYEGLWV
                LRVPSTMVWRPDIVLENNVDGVFEVALYCNVLVSPDGCIYWLPPAIFRSACSISVTYF
                PFDWQNCSLIFQSQTYSTNEIDLQLSQEDGQTIEWIFIDPEAFTENGEWAIQHRPAKM
                LLDPAAPAQEAGHQKVVFYLLIQRKPLFYVINIIAPCVLISSVAILIHFLPAKAGGQK
                CTVAINVLLAQTVFLFLVAKKVPETSQAVPLISKYLTFLLVVTILIVVNAVVVLNVSL
                RSPHTHSMARGVRKVFLRLLPQLLRMHVRPLAPAAVQDTQSRLQNGSSGWSITTGEEV
                ALCLPRSELLFQQWQRQGLVAAALEKLEKGPELGLSQFCGSLKQAAPAIQACVEACNL
                IACARHQQSHFDNGNEEWFLVGRVLDRVCFLAMLSLFICGTAGIFLMAHYNRVPALPF
                PGDPRPYLPSPD"
    gene        complement(77136. .79583)
                /gene="TIGD1"
    mRNA        complement(77136. .79583)
                /gene="TIGD1"
    CDS         complement(77174. .78949)
                /gene="TIGD1"
                /note="Homo sapiens tigger transposable element derived 1
                (TIGD1), mRNA.; H_NH0753F04.4
                This gene was based on gi(22209000)"
                /codon_start=1
                /product="unknown"
                /protein_id="AAY24104.1"
                /db_xref="GI:62988717"
                /translation="MASKCSSERKSRTSLTLNQKLEMIKLSEEGMSKAEIGRRLGLLR
                QTVSQVVNAKEKFLKEVKSATPMNTRMIRKRNSLIADMEKVLVVWIEDQTSRNIPLSQ

  Query Match            26.3%;  Score 526.6;  DB 8;  Length 100020;
  Best Local Similarity  78.7%;  Pred. No. 7.1e-73;
  Matches  767;  Conservative    0;  Mismatches  169;  Indels   39;  Gaps   10;

Qy       1051 ATCACAATAAAGTGAGTCACA-CAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGC 1109
                | ||||| ||||  |||||||  ||   |||||| |||||| ||||||||||||||| ||
Db      77012 AACACAAGAAAGCAAGTCACATGAATGTTTTTGGTTTCCCAATGCATATAAAAGTTATGT 77071

Qy       1110 TTATACTACACTGTAGTCTGTTAAGTGTGCAATAG---TGTTATGTCTAAAAAAACACAT 1166
                ||| |||| ||||||| |    ||| |||||||||   |||    |  || || | ||||
Db      77072 TTACACTATACTGTAGACCAGCAAGAGTGCAATAGCATTGTCTAATAAAACAATATACAT 77131

Qy       1167 ACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC-ATTTGAGCATTCAGT 1225
                |||| ||||| ||||| |||||||| |||||||||||| |  ||| || ||||| ||| ||
Db      77132 ACCTAAATTTAAAAATACTTTATTGCTAAAAAATGCTGATTATCAATCTGAGCCTTCGGT 77191

Qy       1226 GAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---------- 1275
                |||| ||| |||||||||||||||| ||||||  ||    | |||  | |
Db      77192 GAGTCGTACTCTTTTTGCTGGTGGAGGGTCTTGCCTCGAGGTTGATGGTTGCTGACTGGT 77251

Qy       1276 -CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAG 1329
                 ||| |   | |||||||| || |||||||||||||| |||| ||||      ||| | |
Db      77252 CAGGGTGGTGGCTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCCTAAAATAAGACATCGG 77311

Qy       1330 TGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGAT 1389
                ||||| ||      || ||||||||||||||||||||||||||||||||||| ||| || |||
Db      77312 TGAAGCTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTGTAGCATGCGAT 77371

Qy       1390 GCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTC 1448
                ||| ||||||||||||||   |||||||| ||||||| |||||||| ||||||||||
Db      77372 GCTGTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAGTCAATCCTCTC 77431

Qy       1449 AAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCAT 1508
                ||||||||||     || |  | ||||||| || |||||||| ||||| ||||||||||
Db      77432 AAACCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGTTGTCAT 77491

Qy       1509 TTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGA 1568
                |||||||| |||||||  |||||||||| ||||||||||||||||        | |
Db      77492 TTCAACAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGAAACCACTT--- 77548

Qy       1569 ATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCA 1628
                |||||||  |||||||||||  || |||||||||||  ||||||||||||||||||||
Db      77549 -TCTTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATCATGAGATTGCA 77607

Qy       1629 GCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTC 1688
                ||||| |||| |  |||||||     ||  |||| |||||| |||||| ||||| ||||
Db      77608 GCAATTCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCTTTTGCTATTTC 77662

Qy       1689 TACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAG 1748
                ||||||||||    || |||||||| |||||||||||||| |||  |||||| |||||
Db      77663 CACCACATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAGTCATCAATGAG 77722

Qy       1749 GGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGA 1805
                |||  |||||| |||||||||||| |||||||||| ||  | | |||||   ||||||||||
Db      77723 GGTGGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACCTCCTCCCATGA 77782

Qy       1806 ATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTT 1865
                |||| |||||||  |||||||||  ||  |||||||||||||||||| ||||||||||||||
Db      77783 ATCACGAATGTTTTTAATGGCATCTAAAATGGTGAATCCTTTCCAGAAGGTTTTCAATTT 77842

Qy       1866 ACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGG 1925
                ||| | ||||||||||     |||||    |||| || ||| ||| || |||||||||
Db      77843 GCTTTGCCCAGATCCAT-----CAGAGACATCACTATCCATGGCAGCTAGAGCCTTATGA 77897

Qy       1926 AATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGC 1985
                |||||||||||  |||||||| |||||| || |||||||  ||||||||||   ||||
Db      77898 AATGTATTTCTCAAATAATAAGACTTGAAGGTCGAAATTACCCCTTGATCCATGGGCTGC 77957

Qy       1986 AAAATAGATGTTGTG 2000
                | ||| |||||||||
Db      77958 AGAATGGATGTTGTG 77972


RESULT 12
AC012033
LOCUS       AC012033              158276 bp    DNA     linear   HTG 04-SEP-2000
DEFINITION  Homo sapiens clone RP11-91N19, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   AC012033
VERSION     AC012033.20  GI:9966196
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 158276)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
```

```
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158276)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 4, 2000 this sequence version replaced gi:9929521.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HMOR
                Center clone name: RP11-91N19
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 135987 bases at least Q40
                Consensus quality: 145984 bases at least Q30
                Consensus quality: 149762 bases at least Q20
                Estimated insert size: 149948; sum-of-contigs estimation
                Estimated insert size: 249006; agarose-fp estimation
                Quality coverage: 2.8x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    44367: contig of 44367 bp in length
            *    44368    44467: gap of unknown length
            *    44468    73078: contig of 28611 bp in length
            *    73079    73178: gap of unknown length
            *    73179    94206: contig of 21028 bp in length
            *    94207    94306: gap of unknown length
            *    94307   112788: contig of 18482 bp in length
            *   112789   112888: gap of unknown length
            *   112889   120323: contig of 7435 bp in length
            *   120324   120423: gap of unknown length
            *   120424   130440: contig of 10017 bp in length
            *   130441   130540: gap of unknown length
            *   130541   135818: contig of 5278 bp in length
            *   135819   135918: gap of unknown length
            *   135919   141234: contig of 5316 bp in length
            *   141235   141334: gap of unknown length
            *   141335   145232: contig of 3898 bp in length
            *   145233   145332: gap of unknown length
            *   145333   149660: contig of 4328 bp in length
            *   149661   149760: gap of unknown length
            *   149761   150931: contig of 1171 bp in length
            *   150932   151031: gap of unknown length
            *   151032   152757: contig of 1726 bp in length
            *   152758   152857: gap of unknown length
            *   152858   154933: contig of 2076 bp in length
            *   154934   155033: gap of unknown length
            *   155034   156717: contig of 1684 bp in length
            *   156718   156817: gap of unknown length
            *   156818   158276: contig of 1459 bp in length.
FEATURES                Location/Qualifiers
     source             1. .158276
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /clone="RP11-91N19"
     gap                44368. .44467
                        /estimated_length=unknown
     gap                73079. .73178
                        /estimated_length=unknown
     gap                94207. .94306
                        /estimated_length=unknown
     gap                112789. .112888
                        /estimated_length=unknown
     gap                120324. .120423
                        /estimated_length=unknown
     gap                130441. .130540
                        /estimated_length=unknown
     gap                135819. .135918
                        /estimated_length=unknown
     gap                141235. .141334
                        /estimated_length=unknown
     gap                145233. .145332
                        /estimated_length=unknown
     gap                149661. .149760
                        /estimated_length=unknown
     gap                150932. .151031
                        /estimated_length=unknown
     gap                152758. .152857
                        /estimated_length=unknown
     gap                154934. .155033
                        /estimated_length=unknown
     gap                156718. .156817
                        /estimated_length=unknown
ORIGIN

  Query Match           26.3%;  Score 526.6;  DB 14;  Length 158276;
  Best Local Similarity   78.7%;  Pred. No. 6e-73;
  Matches  767;  Conservative    0;  Mismatches  169;  Indels   39;  Gaps   10;

Qy       1051 ATCACAATAAAGTGAGTCACA-CAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGC 1109
                | ||||| |||| |||||||  ||  |||||| ||||||| ||||||||||||||| ||
Db     109769 AACACAAGAAAGCAAGTCACATGAATGTTTTTGGTTTCCCAATGCATATAAAAGTTATGT 109828

Qy       1110 TTATACTACACTGTAGTCTGTTAAGTGTGCAATAG---TGTTATGTCTAAAAAAACACAT 1166
                ||| |||| ||||||| |    ||| |||||||||   |||    | || || | ||||
Db     109829 TTACACTATACTGTAGACCAGCAAGAGTGCAATAGCATTGTCTAATAAAACAATATACAT 109888

Qy       1167 ACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC-ATTTGAGCATTCAGT 1225
                |||| ||||| |||||| ||||||| |||||||||||| |  ||| || ||||| ||| ||
Db     109889 ACCTAAATTTAAAAATACTTTATTGCTAAAAAATGCTGATTATCAATCTGAGCCTTCGGT 109948

Qy       1226 GAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---------- 1275
                |||| ||| |||||||||||||||| ||||||  ||    | |||  | |
Db     109949 GAGTCGTACTCTTTTTGCTGGTGGAGGGTCTTGCCTCGAGGTTGATGGTTGCTGACTGGT 110008

Qy       1276 -CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAG 1329
                ||| |   | |||||||| || |||||||||||||| |||| ||||     ||| | |
Db     110009 CAGGGTGGTGGCTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCCTAAAATAAGACATCGG 110068

Qy       1330 TGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGAT 1389
                ||||| ||      || |||||||||||||||||||||||||||||| |||  || |||
Db     110069 TGAAGCTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTGTAGCATGCGAT 110128
```

```
Qy       1390 GCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTC 1448
             ||| ||||||||||||||   |||||||| ||||||| ||||||||| ||||||||||
Db     110129 GCTGTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAGTCAATCCTCTC 110188

Qy       1449 AAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCAT 1508
             |||||||||     || |  | ||||||| || |||||||| ||||| ||||||||||
Db     110189 AAACCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGTTGTCAT 110248

Qy       1509 TTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGA 1568
             ||||||||| |||||||  |||||||||| ||||||||||||||        | |
Db     110249 TTCAACAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGAAACCACTT--- 110305

Qy       1569 ATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCA 1628
              ||||||| |||||||||||||  || ||||||||||| ||||||||||||||||||||
Db     110306 -TCTTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATCATGAGATTGCA 110364

Qy       1629 GCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTC 1688
             ||||| |||| |  ||||||||     ||  |||| |||||| |||||| ||||| ||||
Db     110365 GCAATTCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCTTTTGCTATTTC 110419

Qy       1689 TACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAG 1748
              ||||||||||   || |||||||| ||||||||||||| |||  ||||||| |||||
Db     110420 CACCACATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAGTCATCAATGAG 110479

Qy       1749 GGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGA 1805
             ||| |||||| |||||||||||| |||||||||| || | | |||||    ||||||||||
Db     110480 GGTGGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACCTCCTCCCATGA 110539

Qy       1806 ATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTT 1865
             |||| ||||||| ||||||||| ||  ||||||||||||||||| ||||||||||||||
Db     110540 ATCACGAATGTTTTTAATGGCATCTAAAATGGTGAATCCTTTCCAGAAGGTTTTCAATTT 110599

Qy       1866 ACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGG 1925
              ||| | ||||||||||     |||||    |||| || ||| ||| || |||||||||
Db     110600 GCTTTGCCCAGATCCAT-----CAGAGACATCACTATCCATGGCAGCTAGAGCCTTATGA 110654

Qy       1926 AATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGC 1985
             ||||||||||  |||||||| |||||| || |||||||| ||||||||||   ||||
Db     110655 AATGTATTTCTCAAATAATAAGACTTGAAGGTCGAAATTACCCCTTGATCCATGGGCTGC 110714

Qy       1986 AAAATAGATGTTGTG 2000
              | ||| |||||||||
Db     110715 AGAATGGATGTTGTG 110729
```

```
RESULT 13
AC074188
LOCUS       AC074188               179816 bp    DNA     linear   HTG 21-FEB-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-341G20, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC074188
VERSION     AC074188.3  GI:13027574
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 179816)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179816)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 21, 2001 this sequence version replaced gi:9802849.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0341G20
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 94%
            Sequencing vector: plasmid; 6%
            Chemistry: Dye-primer ET; 94% of reads
            Chemistry: Dye-terminator Big Dye; 6% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 167169 bases at least Q40
            Consensus quality: 170689 bases at least Q30
            Consensus quality: 172797 bases at least Q20
            Insert size: 186000; agarose-fp
            Insert size: 178216; sum-of-contigs
            Quality coverage: 5.07 in Q20 bases; agarose-fp
            Quality coverage: 5.39 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3838: contig of 3838 bp in length
            *     3839     3938: gap of unknown length
            *     3939     9367: contig of 5429 bp in length
            *     9368     9467: gap of unknown length
            *     9468    15261: contig of 5794 bp in length
            *    15262    15361: gap of unknown length
            *    15362    21615: contig of 6254 bp in length
            *    21616    21715: gap of unknown length
            *    21716    28467: contig of 6752 bp in length
            *    28468    28567: gap of unknown length
            *    28568    36770: contig of 8203 bp in length
            *    36771    36870: gap of unknown length
            *    36871    48944: contig of 12074 bp in length
            *    48945    49044: gap of unknown length
            *    49045    60729: contig of 11685 bp in length
            *    60730    60829: gap of unknown length
            *    60830    74071: contig of 13242 bp in length
            *    74072    74171: gap of unknown length
            *    74172    85808: contig of 11637 bp in length
            *    85809    85908: gap of unknown length
            *    85909   101607: contig of 15699 bp in length
            *   101608   101707: gap of unknown length
            *   101708   120221: contig of 18514 bp in length
            *   120222   120321: gap of unknown length
            *   120322   144619: contig of 24298 bp in length
            *   144620   144719: gap of unknown length
            *   144720   172326: contig of 27607 bp in length
            *   172327   172426: gap of unknown length
            *   172427   174046: contig of 1620 bp in length
            *   174047   174146: gap of unknown length
            *   174147   176137: contig of 1991 bp in length
            *   176138   176237: gap of unknown length
            *   176238   179816: contig of 3579 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179816
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-341G20"
     misc_feature    1. .3838
                     /note="assembly_name:Contig10"
     gap             3839. .3938
                     /estimated_length=unknown
     misc_feature    3939. .9367
                     /note="assembly_name:Contig11"
     gap             9368. .9467
```

```
                     /estimated_length=unknown
     misc_feature    9468. .15261
                     /note="assembly_name:Contig12"
     gap             15262. .15361
                     /estimated_length=unknown
     misc_feature    15362. .21615
                     /note="assembly_name:Contig13"
     gap             21616. .21715
                     /estimated_length=unknown
     misc_feature    21716. .28467
                     /note="assembly_name:Contig14"
     gap             28468. .28567
                     /estimated_length=unknown
     misc_feature    28568. .36770
                     /note="assembly_name:Contig15"
     gap             36771. .36870
                     /estimated_length=unknown
     misc_feature    36871. .48944
                     /note="assembly_name:Contig16"
     gap             48945. .49044
                     /estimated_length=unknown
     misc_feature    49045. .60729
                     /note="assembly_name:Contig17"
     gap             60730. .60829
                     /estimated_length=unknown
     misc_feature    60830. .74071
                     /note="assembly_name:Contig18"
     gap             74072. .74171
                     /estimated_length=unknown
     misc_feature    74172. .85808
                     /note="assembly_name:Contig19"
     gap             85809. .85908
                     /estimated_length=unknown
     misc_feature    85909. .101607
                     /note="assembly_name:Contig20"
     gap             101608. .101707
                     /estimated_length=unknown
     misc_feature    101708. .120221
                     /note="assembly_name:Contig21
                     clone_end:T7
                     vector_side:left"
     gap             120222. .120321
                     /estimated_length=unknown
     misc_feature    120322. .144619
                     /note="assembly_name:Contig22"
     gap             144620. .144719
                     /estimated_length=unknown
     misc_feature    144720. .172326
                     /note="assembly_name:Contig23
                     clone_end:SP6
                     vector_side:right"
     gap             172327. .172426
                     /estimated_length=unknown
     misc_feature    172427. .174046
                     /note="assembly_name:Contig7"
     gap             174047. .174146
                     /estimated_length=unknown
     misc_feature    174147. .176137
                     /note="assembly_name:Contig8"
     gap             176138. .176237
                     /estimated_length=unknown
     misc_feature    176238. .179816
                     /note="assembly_name:Contig9"
ORIGIN

  Query Match           26.3%;  Score 526.6;  DB 14;  Length 179816;
  Best Local Similarity   78.7%;  Pred. No. 5.7e-73;
  Matches  767;  Conservative    0;  Mismatches  169;  Indels   39;  Gaps   10;

Qy       1051 ATCACAATAAAGTGAGTCACA-CAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGC 1109
             | ||||| ||||  |||||||  ||    |||||| |||||| |||||||||||||| ||
Db     149647 AACACAAGAAAGCAAGTCACATGAATGTTTTTGGTTTCCCAATGCATATAAAAGTTATGT 149706

Qy       1110 TTATACTACACTGTAGTCTGTTAAGTGTGCAATAG---TGTTATGTCTAAAAAAACACAT 1166
              ||| |||| ||||||| |    ||| |||||||||   |||    |  || || | ||||
Db     149707 TTACACTATACTGTAGACCAGCAAGAGTGCAATAGCATTGTCTAATAAAACAATATACAT 149766

Qy       1167 ACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC-ATTTGAGCATTCAGT 1225
              |||| ||||| ||||| ||||||| |||||||||||| |  ||| || ||||| ||| ||
Db     149767 ACCTAAATTTAAAAATACTTTATTGCTAAAAAATGCTGATTATCAATCTGAGCCTTCGGT 149826

Qy       1226 GAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---------- 1275
              |||| ||| |||||||||||||||| ||||||  ||    | |||  | |
Db     149827 GAGTCGTACTCTTTTTGCTGGTGGAGGGTCTTGCCTCGAGGTTGATGGTTGCTGACTGGT 149886

Qy       1276 -CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAG 1329
               ||| |   | |||||||| || |||||||||||||| |||| ||||     ||| | |
Db     149887 CAGGGTGGTGGCTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCCTAAAATAAGACATCGG 149946

Qy       1330 TGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGAT 1389
              ||||| ||     || |||||||||||||||||||||||||||||||| |||  || |||
Db     149947 TGAAGCTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTGTAGCATGCGAT 150006

Qy       1390 GCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTC 1448
              ||| ||||||||||||||   |||||||| ||||||| |||||||| ||||||||||
Db     150007 GCTGTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAGTCAATCCTCTC 150066

Qy       1449 AAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCAT 1508
              ||||||||||      || |  | ||||||| || |||||||| ||||| ||||||||||
Db     150067 AAACCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGTTGTCAT 150126

Qy       1509 TTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGA 1568
              |||||||| |||||||  |||||||||| ||||||||||||||        |  |
Db     150127 TTCAACAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGAAACCACTT--- 150183

Qy       1569 ATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCA 1628
              ||||||| ||||||||||||  || |||||||||||| ||||||||||||||||||||
Db     150184 -TCTTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATCATGAGATTGCA 150242

Qy       1629 GCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTC 1688
              ||||| |||| |  ||||||||      ||  |||| |||||| ||||| ||||| ||||
Db     150243 GCAATTCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCTTTTGCTATTTC 150297

Qy       1689 TACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAG 1748
              ||||||||||    || ||||||||| |||||||||||||| |||  ||||||| |||||
Db     150298 CACCACATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAGTCATCAATGAG 150357

Qy       1749 GGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGA 1805
              ||| |||||| |||||||||||| |||||||||| || | | |||||   ||||||||||
Db     150358 GGTGGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACCTCCTCCCATGA 150417

Qy       1806 ATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTT 1865
              |||| ||||||| |||||||||  ||  |||||||||||||||||| ||||||||||||||
Db     150418 ATCACGAATGTTTTTAATGGCATCTAAAATGGTGAATCCTTTCCAGAAGGTTTTCAATTT 150477

Qy       1866 ACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGG 1925
              ||| | ||||||||||     |||||    |||| || ||| ||| || ||||||||||
Db     150478 GCTTTGCCCAGATCCAT-----CAGAGACATCACTATCCATGGCAGCTAGAGCCTTATGA 150532

Qy       1926 AATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGC 1985
              ||||||||||  |||||||| |||||| || |||||||| ||||||||||   ||||
Db     150533 AATGTATTTCTCAAATAATAAGACTTGAAGGTCGAAATTACCCCTTGATCCATGGGCTGC 150592

Qy       1986 AAAATAGATGTTGTG 2000
              | ||| |||||||||
Db     150593 AGAATGGATGTTGTG 150607


RESULT 14
AF307337
LOCUS       AF307337               210301 bp    DNA     linear   PRI 23-APR-2001
DEFINITION  Homo sapiens ALPP, ALPPL2, ALPI, XCE, CHRND, and CHRNG genes,
            complete sequence.
ACCESSION   AF307337
VERSION     AF307337.1  GI:10732831
KEYWORDS    HTG.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 210301)
  AUTHORS   Rump,A., Kasper,G., Hayes,C., Wen,G., Starke,H., Liehr,T.,
            Lehmann,R., Lagemann,D. and Rosenthal,A.
  TITLE     Complex arrangement of genes within a 220-kb region of
            double-duplicated DNA on human 2q37.1
  JOURNAL   Genomics 73 (1), 50-55 (2001)
   PUBMED   11352565
REFERENCE   2  (bases 1 to 210301)
  AUTHORS   Rump,A., Hess,J., Weng,G.P., Galgoczy,P., Schillhabel,M., Wirth,T.
            and Rosenthal,A.
  TITLE     Genomic organisation of the alkaline phosphatase genes on human
            chromosome 2q37.1
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 210301)
  AUTHORS   Rump,A., Hess,J., Weng,G.P., Galgoczy,P., Schillhabel,M., Wirth,T.
            and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2000) Genome Analysis, Molecular Biotechnology,
            Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .210301
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37.1"
     gene            39528. .43621
                     /gene="ALPP"
                     /note="alkaline phosphatase, placental (Regan isozyme);
                     GenBank Accession Number of corresponding cDNA sequence:
                     M14170"
     gene            67687. .71537
                     /gene="ALPPL2"
                     /note="alkaline phosphatase, placental like 2; GenBank
                     Accession Number of corresponding cDNA sequence: X55958"
     unsure          88033. .88060
     gene            117076. .120958
                     /gene="ALPI"
                     /note="alkaline phosphatase, intestinal; GenBank Accession
                     Number of corresponding cDNA: E08258"
     gene            complement(140732. .148527)
                     /gene="XCE"
                     /note="X chromosome controlling element; GenBank Accession
                     Number of corresponding cDNA: Y16187"
     gene            187146. .196428
                     /gene="CHRND"
                     /note="cholinergic receptor, nicotinic, delta polypeptide;
                     GenBank Accession Number of corresponding cDNA: X55019"
     gene            200681. .206639
                     /gene="CHRNG"
                     /note="cholinergic receptor, nicotinic, gamma polypeptide;
                     GenBank Accession Number of corresponding cDNA: NM_005199"
ORIGIN

  Query Match             26.3%;  Score 526.6;  DB 8;  Length 210301;
  Best Local Similarity   78.7%;  Pred. No. 5.4e-73;
  Matches  767;  Conservative    0;  Mismatches  169;  Indels   39;  Gaps   10;

Qy       1051 ATCACAATAAAGTGAGTCACA-CAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGC 1109
Db     208863 AACACAAGAAAGCAAGTCACATGAATGTTTTTGGTTTCCCAATGCATATAAAAGTTATGT 208922

Qy       1110 TTATACTACACTGTAGTCTGTTAAGTGTGCAATAG---TGTTATGTCTAAAAAAACACAT 1166
Db     208923 TTACACTATACTGTAGACCAGCAAGAGTGCAATAGCATTGTCTAATAAAACAATATACAT 208982

Qy       1167 ACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC-ATTTGAGCATTCAGT 1225
```

```
Db     208983 ACCTAAATTTAAAAATACTTTATTGCTAAAAAATGCTGATTATCAATCTGAGCCTTCGGT 209042

Qy       1226 GAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---------- 1275
Db     209043 GAGTCGTACTCTTTTTGCTGGTGGAGGGTCTTGCCTCGAGGTTGATGGTTGCTGACTGGT 209102

Qy       1276 -CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAG 1329
Db     209103 CAGGGTGGTGGCTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCCTAAAATAAGACATCGG 209162

Qy       1330 TGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGAT 1389
Db     209163 TGAAGCTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTGTAGCATGCGAT 209222

Qy       1390 GCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTC 1448
Db     209223 GCTGTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAGTCAATCCTCTC 209282

Qy       1449 AAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCAT 1508
Db     209283 AAACCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGTTGTCAT 209342

Qy       1509 TTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGA 1568
Db     209343 TTCAACAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGAAACCACTT--- 209399

Qy       1569 ATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCA 1628
Db     209400 -TCTTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATCATGAGATTGCA 209458

Qy       1629 GCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTC 1688
Db     209459 GCAATTCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCTTTTGCTATTTC 209513

Qy       1689 TACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAG 1748
Db     209514 CACCACATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAGTCATCAATGAG 209573

Qy       1749 GGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGA 1805
Db     209574 GGTGGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACCTCCTCCCATGA 209633

Qy       1806 ATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTT 1865
Db     209634 ATCACGAATGTTTTTAATGGCATCTAAAATGGTGAATCCTTTCCAGAAGGTTTTCAATTT 209693

Qy       1866 ACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGG 1925
Db     209694 GCTTTGCCCAGATCCAT-----CAGAGACATCACTATCCATGGCAGCTAGAGCCTTATGA 209748

Qy       1926 AATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGC 1985
Db     209749 AATGTATTTCTCAAATAATAAGACTTGAAGGTCGAAATTACCCCTTGATCCATGGGCTGC 209808

Qy       1986 AAAATAGATGTTGTG 2000
Db     209809 AGAATGGATGTTGTG 209823
```

```
RESULT 15
BC035143/c
LOCUS       BC035143                2641 bp    mRNA    linear   PRI 21-JUL-2005
DEFINITION  Homo sapiens cDNA clone MGC:43254 IMAGE:5264460, complete cds.
ACCESSION   BC035143
VERSION     BC035143.1  GI:23243251
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2641)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
```

```
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  CONSRTM   Mammalian Gene Collection Program Team
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2641)
  AUTHORS   .
  CONSRTM   NIH MGC Project
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Bethesda, MD 20892-2590, USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 73 Row: j Column: 13
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 22209000.
FEATURES             Location/Qualifiers
     source          1. .2641
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:43254 IMAGE:5264460"
                     /tissue_type="Brain, hippocampus"
                     /clone_lib="NIH_MGC_95"
                     /lab_host="DH10B"
                     /note="Vector: pBluescriptR"
     CDS             780. .2489
                     /codon_start=1
                     /product="Unknown (protein for MGC:43254)"
                     /protein_id="AAH35143.1"
                     /db_xref="GI:38494046"
                     /translation="MIKLSEEGMSKAEIGRRLGLLRQTVSQVVNAKEKPLKEVKSATP
                     MNTRMIRKRNSLIADMEKVLVVWIEDQTSRNIPLSQSLIQNKALTLFNSMKAERGVEA
                     AEEKFEASRGWFMRFKERSHFHNIKAQGEAASADVEAAASYPEALAKIIDEGGYTKQQ
                     IFNVDETAFYWKKMPSRTFIAREEKSVPGFKASKDRLTLLLGANAAGDFKLKPMLIYH
                     SENPRALKSYTKSTLPVLYKWNSKARMTAHLFTAWFTEYFKPTVETYCSEKKIPFKIL
                     LLIGNAPSHPRALMEIYEEINVIFMPANTTSILQPMDQGVISTFKSYYLRNTFHKALA
                     AMDSDVSDGSGQSKLKTFWKGFTILDAIKNIRDSWEEVKLSTLTGVWKKLIPTLIDDY
                     EGFKTSVEEVSADVVEIAKELELEVEPEDVTELLQSHDKTLTDEELFLMDAQRKWFLE
                     MESTPGEDAVNIVEMTTKDLEYYINLVDKAAAGFERIDSNFERSSTVGKMLSNSIACY
                     REIFHERKSQLMRKASPMSYFRKLPQPPQPSAATTLTSQQPSTSRQDPPPAKRVRLTE
                     GSD"
ORIGIN
```

```
  Query Match           26.2%;  Score 523.4;  DB 8;  Length 2641;
  Best Local Similarity   79.0%;  Pred. No. 8.6e-72;
  Matches  748;  Conservative    0;  Mismatches  161;  Indels   38;  Gaps    9;

Qy       1078 TTTTGGCTTCCCAGTGCATATAAAAGTTTGCTTATACTACACTGTAGTCTGTTAAGTGT 1137
             |||||| |||||| |||||||||||||| || ||| |||| ||||||| |    ||| ||
Db       2623 TTTTGGTTTCCCAATGCATATAAAAGTTATGTTTACACTATACTGTAGACCAGCAAGAGT 2564

Qy       1138 GCAATAG---TGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTA 1194
              |||||||   |||     |  || || | |||||||| ||||| ||||| |||||||| |||
Db       2563 GCAATAGCATTGTCTAATAAAACAATATACATACCTAAATTTAAAAATACTTTATTGCTA 2504

Qy       1195 AAAAATGCTAACAATC-ATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGG 1253
              |||||||| |  ||| || ||||| ||| |||||| ||| |||||||||||||||| ||
Db       2503 AAAAATGCTGATTATCAATCTGAGCCTTCGGTGAGTCGTACTCTTTTTGCTGGTGGAGGG 2444

Qy       1254 TCTTTTCTTATTGATGACTGAT----------CGGGGGTCAGGTGCTGAAGCTTAGGGT 1302
              ||||  ||    | |||  | |           ||| |   | |||||||| || ||||
Db       2443 TCTTGCCTCGAGGTTGATGGTTGCTGACTGGTCAGGGTGGTGGCTGCTGAAGGTTGGGGT 2384

Qy       1303 GGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTC 1357
              |||||||||| |||| ||||      ||| | |||||| ||      || ||||||||||||
Db       2383 GGCTGTGGCAATTTCCTAAAATAAGACATCGGTGAAGCTTTTCGCATGAGTTGACTCTTC 2324

Qy       1358 CTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTA 1417
              |||||||||||||||||||| |||  || |||||| ||||||||||||||   |||||||
Db       2323 CTTTCATGAAAGATTTCTCTGTAGCATGCGATGCTGTTTGATAGCATTTTGCCCACAGTA 2264

Qy       1418 GAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAG 1476
              || ||||||| |||||||| ||||||||||||||||||||||      || |  | |||||
Db       2263 GAGCTTCTTTCAAAATTGGAGTCAATCCTCTCAAACCCTGCTGCTGCTTTATCAACTAAG 2204

Qy       1477 TTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCA 1536
              || || |||||||| ||||| |||||||||||||||||||| |||||||  |||||||||
Db       2203 TTTATGTAATATTCTAAATCCTTTGTTGTCATTTCAACAATGTTCACAGCATCTTCACCA 2144

Qy       1537 GGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTC 1596
              || |||||||||||||||      | |   ||||||| |||||||||||  || ||
Db       2143 GGTGTAGATTCCATCTCAAGAAACCACTT----TCTTTGCGCATCCATAAGAAACAACTC 2088

Qy       1597 CTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCAC 1656
              |||||||||| |||||||||||||||||||||||||| |||| |  ||||||||     |
Db       2087 CTCATCTGTTAAAGTTTTATCATGAGATTGCAGCAATTCAGTTACATCTTCAGG-----C 2033

Qy       1657 TTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATT 1716
              |  |||| |||||| ||||| |||| |||||||||||   || |||||||| |
Db       2032 TCTACTTCTAATTCTAGTTCTTTTGCTATTTCCACCACATCTGCACTTACTTCCTCCACT 1973

Qy       1717 GAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCT 1776
              |||||||||||| ||| ||||||| |||||||| |||||| |||||||||||| ||||
Db       1972 GAAGTCTTGAACCCCTCATAGTCATCAATGAGGGTGGGAATCAACTTCTTCCAAACTCCT 1913

Qy       1777 GTTAATATTTATATTTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGA 1833
              |||||| || | | |||||   |||||||||||||| ||||||| ||||||||| ||  |
Db       1912 GTTAATGTTGACAATTTGACCTCCTCCCATGAATCACGAATGTTTTTAATGGCATCTAAA 1853

Qy       1834 ATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGG 1893
              |||||||||||||||| |||||||||||||| ||| | |||||||||     |||||
Db       1852 ATGGTGAATCCTTTCCAGAAGGTTTTCAATTTGCTTTGCCCAGATCCAT-----CAGAGA 1798

Qy       1894 ATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGA 1953
                 |||| || ||| ||| || |||||||||| ||||||||||  |||||||| |||||
Db       1797 CATCACTATCCATGGCAGCTAGAGCCTTATGAAATGTATTTCTCAAATAATAAGACTTGA 1738

Qy       1954 AAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              | || |||||||| ||||||||||   ||||| ||| |||||||||
Db       1737 AGGTCGAAATTACCCCTTGATCCATGGGCTGCAGAATGGATGTTGTG 1691
```

```
RESULT 16
AL596257/c
LOCUS       AL596257              60169 bp    DNA     linear   PRI 18-MAY-2005
```

```
DEFINITION  Human DNA sequence from clone RP11-251B9 on chromosome 1, complete
            sequence.
ACCESSION   AL596257 AC027619
VERSION     AL596257.5  GI:16973150
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 60169)
  AUTHORS   Lad,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:16304759.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-251B9 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            Draft Sequence Produced by Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            http://genome.wustl.edu/gsc/index.shtml
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .60169
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-251B9"
                     /clone_lib="RPCI-11.1"
     misc_feature    2000
                     /note="Clone_right_end: RP4-614N24"
     misc_feature    58170
                     /note="Clone_left_end: RP4-655C4"
ORIGIN

  Query Match             26.1%;  Score 522.2;  DB 8;  Length 60169;
  Best Local Similarity   75.8%;  Pred. No. 4.1e-72;
  Matches  798;  Conservative    0;  Mismatches  198;  Indels   57;  Gaps   10;

Qy        976 GTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAAT 1035
              |||||| ||| ||| ||||||| |||| |||   ||||| || ||  ||||||  |||||
Db      48252 GTACAGACATGCCTCGGAGATATTGTGAGTTCAATTCCAGACTACTGCAATAAAGCAAAT 48193

Qy       1036 ATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCA 1095
              ||                  ||||||||  | |||||  | | ||||||  || ||||||||
Db      48192 ATTG--------------CAATAAAGCAATTCACATGAATTTTTTGGTTTTCCAGTGCA 48148

Qy       1096 TATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTA 1155
              |||||| || || ||| |||| |||||||||| |||| |||||||||| |||||||||
Db      48147 CATAAAACTTATGTTTACACTATACTGTAGTCTATTAACTGTGCAATAGTATTATGTCTA 48088

Qy       1156 AAAAAA---CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212
              |||| |    | |||| ||||||| |||| |   || |   |||||||||| ||||||
Db      48087 AAAACAATGGATATACTTTAATTTAAAAAACTTTATTGCTGAAAAAATGCTAGCAATCAG 48028

Qy       1213 TTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTG-----GAAGGTCTTTTCTTATTGA 1267
              ||||| ||||| |||||||||||||||||||||||     || ||||||  |||  ||
Db      48027 CTGAGCCTTCAGCGAGTTGTAATCTTTTTGCTGGTGGAGTAGAGGGTCTTGCCTTGATGT 47968

Qy       1268 TGACTGATCGGGGGTCA-----------GGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              ||| | |   |  |||           | ||| |||| ||  |||| |||||||| |||
Db      47967 TGATGGCTGCTGACTCATCAGGGTAGTGGTTGCGGAAGGTTGTGGTGACTGTGGCAATTT 47908

Qy       1317 CTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              ||||||     |||||||  ||| ||||   |||   |||| ||||||| ||||||||||
Db      47907 CTTAAAGTAAGACAACAGGAAAGTTTGCTGCATCCAGTGACACTTCCTTCCATGAAAGAT 47848

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAAACTTCTTTGAAA 1431
              |||||| |||  |   ||||| |||||||||||||||  |||||||||||||||||| |||
Db      47847 TTCTCTGTAGGATACCATGCTGTTTGATAGCATTTTACCCACAGTAGAACTTCTTTCAAA 47788

Qy       1432 ATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              ||| || ||| |||||||||||||||||      || |  | || |||| | || |||| |
Db      47787 ATTAGACTCAGTCCTCTCAAACCCTGCTGCTGCTTTATCAACTGAGTTTACATGATATCC 47728

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
              | |||||| ||| |||||||||||||| |||||||  |||||||||||||||||||| ||
Db      47727 TAAATCCACTGTCGTCATTTCAACAATGTTCACAGCATCTTCACCAGGAGTAGATTCTAT 47668

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              || |        |  |    |||||||||||||||||||| || |||||||| ||| |||
Db      47667 CTTAAGAAACCACTT----TCTTTGCTCATCCATAAGAAGCAACTCCTCATCCGTTAAAG 47612

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              |||||||||||||||||| ||| ||||||  ||||||||      || ||||| ||||||
Db      47611 TTTTATCATGAGATTGCAGTAATTCAGTCACATCTTCAGG-----CTCCACTTCTAATTC 47557

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
              |||||||||||| |||| |||||||||| ||| ||||||  |   || |||||||||
Db      47556 TAGTTCTCTTGCTATTTCCACCACATCTGCAGTTACTTCCTGTACACAAATCTTGAACCC 47497

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              ||| || | |||| ||||| |||||| | |||||||||||||||| |||  |  || ||||
Db      47496 CTCAAAATAATCCCTGAGGATTGGAACCAACTTCTTCCAAATTCTTGTACAAGTTGATAT 47437

Qy       1791 TTTGAC---CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
              ||||||   || |||||||||||||||||||||||||||| || ||||| ||||||||||
Db      47436 TTTGACCTTCTGCCATGAATCATGAATGTTCTTAATGGCATCTAGAATGATGAATCCTTT 47377

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              ||| |||||||||||||| ||| | |||||||||      ||| || |  |||| || |||
Db      47376 CCAGAAGGTTTTCAATTTTCTTTGCCCAGATCCA-----CCAAAGAAGTCACTATCTATG 47322

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
              ||| | ||||||||    |||||| |||| ||| |||||  |||||||||  |||||| |
Db      47321 GCAGCTCTAGCCTTACAAAATGTACTTCTTAAATTATAAGACTTGAAAGTCAAAATTATT 47262

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              ||| |||||||   ||||| | | |||||||||
Db      47261 CCTCGATCCATAGGCTGCAGATTGGATGTTGTG 47229

RESULT 17
AC108017
LOCUS       AC108017               51343 bp    DNA     linear   PRI 13-MAY-2005
DEFINITION  Homo sapiens BAC clone CTD-2005A22 from 4, complete sequence.
ACCESSION   AC108017
VERSION     AC108017.4  GI:19526131
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 51343)
  AUTHORS   Kozlowicz,A., Dignan,G. and Trani,L.
  TITLE     The sequence of Homo sapiens BAC clone CTD-2005A22
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 51343)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 51343)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 51343)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 51343)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 51343)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 16, 2002 this sequence version replaced gi:18873881.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_MS2005A22
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted;
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            Clone CTD-2005A22 is from a release of the human BAC library CTD.
            The library contains cloned DNA from human sperm. See:  Shizuya et
            al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al.,
            Genomics 34:213-8 (1996). The clone is available from Research
            Genetics, Inc. (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:

            The clone sequenced to the left is RP11-719M18, 2000 bp overlap;
            the clone sequenced to the right is RP11-575B4, 2000 bp overlap.
            Actual start of this clone is at base position 120807 of
            RP11-719M18; actual end is at base position 7921 of RP11-575B4.

            Polymorphisms have been identified between AC108017 and AC019341.
FEATURES            Location/Qualifiers
     source         1. .51343
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="4"
                    /clone="CTD-2005A22"
                    /clone_lib="CTD"
     gene           7452. .47527
                    /gene="LGR7"
     mRNA           join(7452. .7550,13377. .13482,19119. .19190,22023. .22094,
                    26181. .26252,26342. .26413,31182. .31256,40871. .40942,
                    42689. .42760,47456. .47527)
                    /gene="LGR7"
     CDS            join(<7452. .7550,13377. .13482,19119. .19190,22023. .22094,
                    26181. .26252,26342. .26413,31182. .31256,40871. .40942,
                    42689. .42760,47456. .>47527)
                    /gene="LGR7"
                    /note="Homo sapiens leucine-rich repeat-containing G
                    protein-coupled receptor 7 (LGR7), mRNA.; H_MS2005A22.1
                    This gene was based on gi(11056007)
                    Continued from H_NH0719M18.1
                    Continues as H_NH0575B04.1"
                    /codon_start=3
                    /product="unknown"
                    /protein_id="AAY41039.1"
                    /db_xref="GI:63995606"
                    /translation="DNNGWSLQFDKYFASYYKMTSQYPFEAETPECLVGSVPVQCLCQ
                    GLELDCDETNLRAVPSVSSNVTAMSLQWNLIRKLPPDCFKNYHDLQKLYLQNNKITSI
                    SIYAFRGLNSLTKLYLSHNRITFLKPGVFEDLHRLEWLIIEDNHLSRISPPTFYGLNS
                    LILLVLMNNVLTRLPDKPLCQHMPRLHWLDLEGNHIHNLRNLTFISCSNLTVLVMRKN
                    KINHLNENTFAPLQKLDELDLGSNKIENLPPLIFKDLKELSQ"
ORIGIN

  Query Match            25.9%;  Score 518.8;  DB 8;  Length 51343;
  Best Local Similarity  75.1%;  Pred. No. 1.5e-71;
  Matches  787;  Conservative    0;  Mismatches  207;  Indels   54;  Gaps    9;

Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
              |||||||| ||||  | |||| ||||||| | |||||| |||||   |||||  |||||
Db      26996 TACAGGCAGACCTCAGTGATATTGTGGGTGTTGTTCCAGACCACTGTAATAAAGCAAATG 27055

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCA---GTG 1093
                          ||||||||||  |||||||||| | |||| | || ||    |
Db      27056 ---------------TCACAATAAAGCAAGTCACACAAATTTTTTTGGTTTCCCAGTGCA 27100

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
              ||||  |||||| || ||| |||| ||| || ||  ||||||||||||||   ||||||
Db      27101 CATAATAAAGTTATGTTTACACTATACTATATTCAATTAAGTGTGCAATAACATTATGTT 27160

Qy       1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
              ||||||    |||||||||||||||||  ||||||| |||||||  ||||||||||||
Db      27161 TAAAAATGTGCATACCTTAATTTTAAAACTCTTTATTGCTAAAAAGAGCTAACAATCATC 27220

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTG 1273
              ||||  ||||| ||||  |||||||||||||||| | |||||  | | | || |||| |
Db      27221 TGAGTTTTCAGCGAGTCATAATCTTTTTGCTGGTGAAGGGTCTCATGTCAATGTTGACAG 27280

Qy       1274 ATCGGGGGTCA----------GGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA 1322
                |   | | |          | |||||||| | |||||||||||||| |||||||||
Db      27281 CTGCTGACTAACCACAGTAGTGGTTGCTGAAGCCTGGGGTGGCTGTGGCAATTTCTTAAA 27340

Qy       1323 -----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCT 1377
                  | |||| ||| | |||| | ||   |||||| ||  ||||||||||||||||||
Db      27341 ATAAGATAACAATGAGGTTTGCTACATTGATTGACTTTTAATTTCATGAAAGATTTCTCT 27400

Qy       1378 CTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA 1437
```

```
                 | || ||  ||||| |||| ||||||||||  ||||  |||| |||||| |||||||  |
Db       27401 ATGGTATGCAATGCTATTTGCTAGCATTTTACCCACAAGAGAATTTCTTTCAAAATTGAA 27460

Qy        1438 -TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATC 1496
               ||||| |||||||| ||| |       || |  |  |||||| || |||||||| |||
Db       27461 GTCAATACTCTCAAATCCTCCCAATGCTTTATCAAATAAGTTTATGTAATATTCTAAATA 27520

Qy        1497 CATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATT 1556
               | ||||||||||||||||| | || ||||  |||| ||||||||||||||||||||||
Db       27521 CTTTGTTGTCATTTCAACAGTGTTTACAGCATCTTTACCAGGAGTAGATTCCATCTCA-- 27578

Qy        1557 TCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTAT 1616
                  ||     |||||| |||||||||||||| || ||||||||||||||||||||||
Db       27579 --AGAAGCCACTTTCTTTGTTCATCCATAAGAAGCAACTCCTCATCTGTTCAAGTTTTAT 27636

Qy        1617 CATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTC 1676
               ||||||||||||||||| ||||||  ||||||||     || ||||| ||||||  |||||
Db       27637 CATGAGATTGCAGCAATTCAGTCACCTCTTCAGG-----CTCCACTTCTAATTCTAGTTC 27691

Qy        1677 TCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAA 1736
               ||||||| |||||||  | |||   ||| |||| |||| ||||||||||||||  || ||
Db       27692 TCTTGCTATTTCTACTTCTTCTACAGTTACTTCTTCCACTGAAGTCTTGAACCCTTCAAA 27751

Qy        1737 GTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC 1796
               |||| |||||||||||||| || |||||||||||| |||||||||| |  ||||||||||
Db       27752 GTCACCCATGAGGGTTGGAGTCAACTTCTTCCAAACTCCTGTTAATGTGAATATTTTGAC 27811

Qy        1797 C----TCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAA 1852
               |    |  |||||||||  |||||||||||||||| || |||||||||| ||||| | |
Db       27812 CTTGTTTTCATGAATCACAAATGTTCTTAATGGCATCTAGAATGGTGAATGCTTTCTAGA 27871

Qy        1853 AGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGT 1912
               ||||||| |||| |||| | ||| |||         |||||||  |||| ||  || ||
Db       27872 AGGTTTTTAATTGACTTTGCCCAAATC------GTCAGAGGAATCACTATCTGTGGCATC 27925

Qy        1913 TATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTG 1972
                 ||||||||   ||||||  |||| ||||||||| |||||||||| ||||||    |||
Db       27926 AGTAGCCTTACAAAATGTACATCTTAAATAATAAGACTTGAAAGTTAAAATTAGCTTTTG 27985

Qy        1973 ATCCATTTTCTGCAAAATAGATGTTGTG 2000
               ||||||   ||| | ||| ||||| |||
Db       27986 ATCCATGGGCTGTAGAATGGATGTCGTG 28013
```

```
RESULT 18
AC158447
LOCUS       AC158447               135818 bp    DNA     linear   HTG 12-APR-2005
DEFINITION  Loxodonta africana clone VMRC15-156A5, WORKING DRAFT SEQUENCE, 4
            ordered pieces.
ACCESSION   AC158447
VERSION     AC158447.2  GI:62510111
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Loxodonta africana (African savanna elephant)
  ORGANISM  Loxodonta africana
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Afrotheria; Proboscidea; Elephantidae;
            Loxodonta.
REFERENCE   1  (bases 1 to 135818)
  AUTHORS   Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W., Boakye,A.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,H.,
            Engle,J., Fuksenko,T., Gestole,M., Greene,A., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hunter,G.,
            Hurle,B., Idol,J.R., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
            Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H.,
            Masiello,C., Maskeri,B., McDowell,J., Mojidi,H.A., Mullikin,J.C.,
            Oestreicher,J.S., Park,M., Portnoy,M.E., Prasad,A., Puri,O.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Stephen,E., Taye,A., Thomas,J.W., Thomas,P.J.,
            Tsipouri,V., Ung,L., Vogt,J.L., Wetherby,K.D., Young,A. and
            Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 135818)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-2005) NIH Intramural Sequencing Center, 5625
            Fishers Lane, Rockville, MD 20852, USA
REFERENCE   3  (bases 1 to 135818)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2005) NIH Intramural Sequencing Center, 5625
            Fishers Lane, Rockville, MD 20852, USA
COMMENT     On Apr 12, 2005 this sequence version replaced gi:61326250.
            ------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ------------- Project Information
                 Center project name: jlu
                 Center clone name: 156A05

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is generally based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 135058 bases at least Q40
                 Consensus quality: 135283 bases at least Q30
                 Consensus quality: 135424 bases at least Q20
                 Insert size: 142000; agarose-fp
                 Insert size: 135518; sum-of-contigs
                 Quality coverage: 9.47x in Q20 bases; agarose-fp
                 Quality coverage: 9.92x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    30739: contig of 30739 bp in length
            *    30740    30839: gap of unknown length
            *    30840   114090: contig of 83251 bp in length
            *   114091   114190: gap of unknown length
            *   114191   129767: contig of 15577 bp in length
            *   129768   129867: gap of unknown length
            *   129868   135818: contig of 5951 bp in length.
FEATURES             Location/Qualifiers
     source          1. .135818
                     /organism="Loxodonta africana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9785"
                     /clone="VMRC15-156A5"
                     /clone_lib="VMRC15"
                     /note="BAC resource: http://bacpac.chori.org/"
     misc_feature    1. .30739
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     gap             30740. .30839
                     /estimated_length=unknown
```

```
     misc_feature    30840. .114090
                     /note="assembly_fragment"
     misc_feature    70837. .135818
                     /note="clone overlaps with GenBank Accession Number
                     AC157635 clone VMRC15-551I19 (center project name jlt)"
     gap             114091. .114190
                     /estimated_length=unknown
     misc_feature    114191. .129767
                     /note="assembly_fragment"
     gap             129768. .129867
                     /estimated_length=unknown
     misc_feature    129868. .135818
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             25.8%;  Score 516.8;  DB 14;  Length 135818;
  Best Local Similarity   76.2%;  Pred. No. 2.1e-71;
  Matches  802;  Conservative    0;  Mismatches  187;  Indels   63;  Gaps   11;

Qy        970 TAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAAT 1029
              | ||  ||||||||||||||||||||| || || || || |||| ||||||||||||
Db     107981 TTGAATATACAGGCATACCTTGGAGATATTGCGGATTCGGCTCCAGACCACCACAATAAA 108040

Qy       1030 ACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCC 1089
               |||||||   |                    |||||||    | ||||||  || ||
Db     108041 GCAAATATCGCA--------------------AGTCACATGGATTTTTTGGTTTACC 108077

Qy       1090 AGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTA 1149
              ||| ||| |||||||| || |||| ||| | |||||||| ||||||||||||||  |||
Db     108078 AGTACATGTAAAAGTTATGTTTATGCTATAATGTAGTCTATTAAGTGTGCAATAGCATTA 108137

Qy       1150 TGTCTAAAAAAAC-----ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTA 1204
              |||||  |||| |     ||||||||||||  ||||| |||||||  |||||||||||||
Db     108138 TGTCTTGAAAACCAATGTACATACCTTAATTAAAAAATACTTTATTGCTAAAAAATGCTA 108197

Qy       1205 ACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTAT 1264
              || ||||| ||||| ||  |  ||| |||||||||||||||||| || |||  |||
Db     108198 ACCATCATCTGAGCCTTGTGCAAGTGGTAATCTTTTTGCTGGTGGAGGGGCTTGCCTTGA 108257

Qy       1265 TGATGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAG 1313
              || |||  | |            ||| |       |||||| || ||||||||||||||
Db     108258 TGTTGATAGCTGCTGACTGATCAGGGTGGTGATCACTGAAGGTTGGGGTGGCTGTGGCAG 108317

Qy       1314 TTTCTTAAAACA--ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              |||||||||| |  ||| |   | ||||   ||   ||||   ||||||||||||||||
Db     108318 TTTCTTAAAATAAGACAATTCTGTTTGCCCCATTGATTGATGTTTCCTTTCATGAAAGAT 108377

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
              |||||| |||  |  |||||| ||||||||||||||    |||| | |||||||| |||
Db     108378 TTCTCTGTAGCATCCGATGCTATTTGATAGCATTTTTCCTACAGCAAAACTTCTTTCAAA 108437

Qy       1432 ATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              |||||| |||||||||||||| ||||||  | |||||  || ||||||| || ||||||||
Db     108438 ATTGGAGTCAATCCTCTCAAGCCCTGCCATTCTTTATTAA-CTAAGTTTATGTAATATTC 108496

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
              | ||||| ||||||||||||||||  |||||||  |||||||||||||||||||||||
Db     108497 TAAATCCTTTGTTGTCATTTCAACAACATTCACAGCATCTTCACCAGGAGTAGATTCCAT 108556

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              ||||       | ||    |||||||||||||||||||| || ||| |||||||| |||
Db     108557 CTCAAGAAACCACAT----TCTTTGCTCATCCATAAGAAGCAACTCCCCATCTGTTAAAG 108612

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              |||||||||||||||||||||| || |||  ||||| ||    ||  |||| ||||||
Db     108613 TTTTATCATGAGATTGCAGCAATTCACTCACATCTTCTGG-----CTCTACTTCTAATTC 108667

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
                |||||||||| |||| ||||  ||||| ||| ||||| ||| ||||||||||| ||
Db     108668 TCCTTCTCTTGCTATTTCCACCATGTCTGTAGTTACTTCCCCCACTGAAGTCTTGATCCC 108727

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              ||| |||||||||||| |||| || |||| |||||||||||| ||||||| || || | |
Db     108728 CTCAAAGTCATCCATAAGGGCTGAAATCAACTTCTTCCAAACTCCTGTTCATGTTGAGGT 108787

Qy       1791 TTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
              |||||   |||||||||||  |||||||||| |||||| || ||||| |||||||||
Db     108788 TTTGACCTCCTCCCATGAATGGTGAATGTTCTTGATGGCATCTAGAATGTTGAATCCTTT 108847

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              || |||||||||| ||||||| | ||||||||||     |||||||  |||  ||  ||
Db     108848 TCAGAAGGTTTTCAGTTTACTTTGCCCAGATCCAT-----CAGAGGAATCACCATCTGTG 108902

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
              ||| ||||||||||    |||||||||||     |||||| ||| ||||||||||||||
Db     108903 GCAGCTATAGCCTTACAAAATGTATTTCTT---AAATAAGACTTAAAAGTTGAAATTACT 108959

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGT 1999
              |||||| ||||   ||||| ||| |||  |||
Db     108960 CCTTGACCCATGGGCTGCAGAATGGATAATGT 108991


RESULT 19
AC157635
LOCUS       AC157635               141510 bp    DNA     linear   HTG 01-APR-2005
DEFINITION  Loxodonta africana clone VMRC15-551I19, WORKING DRAFT SEQUENCE, 6
            ordered pieces.
ACCESSION   AC157635
VERSION     AC157635.2  GI:62123027
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Loxodonta africana (African savanna elephant)
  ORGANISM  Loxodonta africana
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Afrotheria; Proboscidea; Elephantidae;
            Loxodonta.
REFERENCE   1  (bases 1 to 141510)
  AUTHORS   Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W., Boakye,A.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,H.,
            Engle,J., Fuksenko,T., Gestole,M., Greene,A., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hunter,G.,
            Hurle,B., Idol,J.R., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
            Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H.,
            Masiello,C., Maskeri,B., McDowell,J., Mojidi,H.A., Mullikin,J.C.,
            Oestreicher,J.S., Park,M., Portnoy,M.E., Prasad,A., Puri,O.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Stephen,E., Taye,A., Thomas,J.W., Thomas,P.J.,
            Tsipouri,V., Ung,L., Vogt,J.L., Wetherby,K.D., Young,A. and
            Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141510)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2005) NIH Intramural Sequencing Center, 5625
            Fishers Lane, Rockville, MD 20852, USA
REFERENCE   3  (bases 1 to 141510)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2005) NIH Intramural Sequencing Center, 5625
            Fishers Lane, Rockville, MD 20852, USA
COMMENT     On Apr 1, 2005 this sequence version replaced gi:60223201.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: jlt
                 Center clone name: 551I19

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
```

```
          data from individual subclones, overlaps with neighboring
          clones, alignment with available reference sequence (e.g.,
          human), and/or confirmation by PCR testing. In addition,
          the sequence assembly is generally based on at least 8X average
          coverage in Q20 bases and has been reviewed to rule out
          gross misassemblies, the low-quality ends of sequence
          contigs have been trimmed away, and each base is associated
          with a Phrap-derived quality score.
          -------------- Summary Statistics
               Sequencing vector: plasmid; n/a; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.990319
               Consensus quality: 140629 bases at least Q40
               Consensus quality: 140781 bases at least Q30
               Consensus quality: 140889 bases at least Q20
               Insert size: 148000; agarose-fp
               Insert size: 141010; sum-of-contigs
               Quality coverage: 9.15x in Q20 bases; agarose-fp
               Quality coverage: 9.61x in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 6 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1    42915: contig of 42915 bp in length
          *    42916    43015: gap of unknown length
          *    43016    58130: contig of 15115 bp in length
          *    58131    58230: gap of unknown length
          *    58231    88548: contig of 30318 bp in length
          *    88549    88648: gap of unknown length
          *    88649   106602: contig of 17954 bp in length
          *   106603   106702: gap of unknown length
          *   106703   114237: contig of 7535 bp in length
          *   114238   114337: gap of unknown length
          *   114338   141510: contig of 27173 bp in length.
FEATURES             Location/Qualifiers
     source          1. .141510
                     /organism="Loxodonta africana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9785"
                     /clone="VMRC15-551I19"
                     /clone_lib="VMRC15"
                     /note="BAC resource: http://bacpac.chori.org/"
     misc_feature    1. .64115
                     /note="clone overlaps with GenBank Accession Number
                     AC158447 clone VMRC15-156A5 (center project name jlu)"
     misc_feature    1. .42915
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             42916. .43015
                     /estimated_length=unknown
     misc_feature    43016. .58130
                     /note="assembly_fragment"
     gap             58131. .58230
                     /estimated_length=unknown
     misc_feature    58231. .88548
                     /note="assembly_fragment"
     gap             88549. .88648
                     /estimated_length=unknown
     misc_feature    88649. .106602
                     /note="assembly_fragment"
     misc_feature    103152. .141510
                     /note="clone overlaps with GenBank Accession Number
                     AC157633 clone VMRC15-499C6 (center project name jls)"
     gap             106603. .106702
                     /estimated_length=unknown
     misc_feature    106703. .114237
                     /note="assembly_fragment"
     gap             114238. .114337
                     /estimated_length=unknown
     misc_feature    114338. .141510
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match            25.8%;  Score 516.8;  DB 14;  Length 141510;
  Best Local Similarity  76.2%;  Pred. No. 2e-71;
  Matches  802;  Conservative    0;  Mismatches  187;  Indels   63;  Gaps   11;

Qy       970 TAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAAT 1029
             | ||   |||||||||||||||||||||| || || || || |||| ||||||||||||
Db     37144 TTGAATATACAGGCATACCTTGGAGATATTGCGGATTCGGCTCCAGACCACCACAATAAA 37203

Qy      1030 ACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCC 1089
              |||||||   |                        |||||||    | |||||| || ||
Db     37204 GCAAATATCGCA----------------------AGTCACATGGATTTTTTGGTTTACC 37240

Qy      1090 AGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTA 1149
             ||| ||| |||||||| || |||| ||| | |||||||| ||||||||||||||||  |||
Db     37241 AGTACATGTAAAAGTTATGTTTATGCTATAATGTAGTCTATTAAGTGTGCAATAGCATTA 37300

Qy      1150 TGTCTAAAAAAAC-----ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTA 1204
             |||||  |||| |     |||||||||||||  ||||| |||||||||||||||||||||
Db     37301 TGTCTTGAAAACCAATGTACATACCTTAATTAAAAAATACTTTATTGCTAAAAAATGCTA 37360

Qy      1205 ACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTAT 1264
             || ||||| ||||| ||   |  ||| |||||||||||||||||||| || |||  |||
Db     37361 ACCATCATCTGAGCCTTGTGCAAGTGGTAATCTTTTTGCTGGTGGAGGGGCTTGCCTTGA 37420

Qy      1265 TGATGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAG 1313
             || |||  | |           ||| |      |||||| || ||||||||||||||||
Db     37421 TGTTGATAGCTGCTGACTGATCAGGGTGGTGATCACTGAAGGTTGGGGTGGCTGTGGCAG 37480

Qy      1314 TTTCTTAAAACA--ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
             |||||||||| |  ||| |   | ||||   ||   ||||   ||||||||||||||||||
Db     37481 TTTCTTAAAATAAGACAATTCTGTTTGCCCCATTGATTGATGTTTCCTTTCATGAAAGAT 37540

Qy      1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
             |||||| |||  |  |||||| |||||||||||||     |||| | |||||||| |||
Db     37541 TTCTCTGTAGCATCCGATGCTATTTGATAGCATTTTTCCTACAGCAAAACTTCTTTCAAA 37600

Qy      1432 ATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
             |||||| |||||||||||| ||||||  | |||||  || ||||||| || ||||||||
Db     37601 ATTGGAGTCAATCCTCTCAAGCCCTGCCATTCTTTATTAA-CTAAGTTTATGTAATATTC 37659

Qy      1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
             | ||||| ||||||||||||||||||  |||||||  ||||||||||||||||||||||||
Db     37660 TAAATCCTTTGTTGTCATTTCAACAACATTCACAGCATCTTCACCAGGAGTAGATTCCAT 37719

Qy      1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
             ||||       | ||    |||||||||||||||||||||| || ||| |||||||| |||
Db     37720 CTCAAGAAACCACAT----TCTTTGCTCATCCATAAGAAGCAACTCCCCATCTGTTAAAG 37775

Qy      1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
             |||||||||||||||||||||| || |||  ||||| ||     ||  |||| ||||||
Db     37776 TTTTATCATGAGATTGCAGCAATTCACTCACATCTTCTGG-----CTCTACTTCTAATTC 37830

Qy      1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
               |||||||||| |||| ||||  ||||| ||| ||||| ||| |||||||||||  ||
Db     37831 TCCTTCTCTTGCTATTTCCACCATGTCTGTAGTTACTTCCCCCACTGAAGTCTTGATCCC 37890

Qy      1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
             ||| |||||||||| |||| || |||| |||||||||||| ||||||| || || | |
Db     37891 CTCAAAGTCATCCATAAGGGCTGAAATCAACTTCTTCCAAACTCCTGTTCATGTTGAGGT 37950

Qy      1791 TTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
             |||||   |||||||||||  |||||||||| |||||| || ||||| ||||||||||
Db     37951 TTTGACCTCCTCCCATGAATGGTGAATGTTCTTGATGGCATCTAGAATGTTGAATCCTTT 38010
```

```
Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
Db      38011 TCAGAAGGTTTTCAGTTTACTTTGCCCAGATCCAT-----CAGAGGAATCACCATCTGTG 38065

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
Db      38066 GCAGCTATAGCCTTACAAAATGTATTTCTT---AAATAAGACTTAAAAGTTGAAATTACT 38122

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGT 1999
Db      38123 CCTTGACCCATGGGCTGCAGAATGGATAATGT 38154


RESULT 20
AX833395
LOCUS       AX833395               3895 bp    DNA     linear   PAT 15-DEC-2003
DEFINITION  Sequence 519 from Patent EP1347046.
ACCESSION   AX833395
VERSION     AX833395.1  GI:39919530
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
  TITLE     Full-length cDNA sequences
  JOURNAL   Patent: EP 1347046-A 519 24-SEP-2003;
            Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1. .3895
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             25.8%;  Score 516;  DB 6;  Length 3895;
  Best Local Similarity   77.5%;  Pred. No. 1e-70;
  Matches  741;  Conservative    0;  Mismatches  175;  Indels   40;  Gaps    8;

Qy       1071 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 1130
Db       1362 ATAAATTTTTTGGTTTCCCAGTACATGTAAAAGTTATAATTACACTATACTATAGTCTAT 1421

Qy       1131 TAAGTGTGCAATAGTGTTATGTCTAAAAAAA------CACATACCTTAATTTTAAAATGC 1184
Db       1422 TAAGGGTGCAATAGCATTATGTCTGAAAAAAACAATGGTCATGCCTTAATTTAAAAATAC 1481

Qy       1185 TTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCT 1244
Db       1482 TTTATTACTAAAAAATGCTAGCAATCATCTGAACCTTCAGTGAGTCATAATATTTTTGCT 1541

Qy       1245 GGTGGAAGGTCTTTTCTTATTGATGACTGAT----------CGGGGGTCAGGTGCTGAA 1293
Db       1542 AGTGGAGGGCCTTGCCTGGATGTTGATGGCTGCTGACTGATCAGGATGGTGGTTGCTGAA 1601

Qy       1294 GCTTAGGGTGGCTGTGGCAGTTTCT-----TAAAACAACAGTGAAGATTGCAATATCAGT 1348
Db       1602 GGTTAGGGTGGCTGCGGCAATTTCTTAACATAAGACAACAGTGAAGTTTCCTGCATTGAT 1661

Qy       1349 TGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTA 1408
Db       1662 TGACTCTTCCTTTCATGAAATATTTCTCTATAGCATGTAATGCTGTTTGATAGCATTTTA 1721

Qy       1409 TGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAA 1467
Db       1722 CCCACAGTAGAATTTATTTCAGAATTGAAGTCAGTCCTCTTAAACCCTGTCACTTCTTTA 1781

Qy       1468 CAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTG 1527
Db       1782 TTAACTAAGTTCACGCAATATTCTAAATCCTTTGTTGTCATTTCAACAATGTTCACAACA 1841

Qy       1528 TCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAG 1587
Db       1842 TCTTCACCAGGGATAGATTCTGTCTCAAGAAACCACTT----TCTTTGTTCATTCGTAAG 1897

Qy       1588 AAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTC 1647
Db       1898 AAGCAACCCCTCATCTGTTCAAGTTTTCTCATGAGATTGGAGCAATTCAGTCACATTTTC 1957

Qy       1648 AGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCT 1707
Db       1958 AGG-----CTGCATTTTTAATTCTAGTTCTCTTGCTATTTCTACCACATCTGCAGTTACT 2012

Qy       1708 TCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTC 1767
Db       2013 TCCTTCACTGAAGGCTTGAACCCCTCAAAGTCACCCATGAGGGTTGGAATCAACTTCTTC 2072

Qy       1768 CAAATTCCTGTTAATATTTATATTTTGACCTCC---CATGAATCATGAATGTTCTTAATG 1824
Db       2073 CAAATTCCTGTTAATGTTGATGTTTTGACCTCCTCTCATGAATGATGAATGTTTTTAATG 2132

Qy       1825 GCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCC 1884
Db       2133 ACATCTAAAATGGTCAGTCCTTTCCAGAAGGTTTTCAATTTACTTTGCCCCAATCCAT-- 2190

Qy       1885 ATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAAT 1944
Db       2191 ---CAGAGGAATCACTGTCTATGGCAACTGTAGGCTTATAAAAGGTATTTCTTAAATAAT 2247

Qy       1945 AAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
Db       2248 ATGTCTTGAAAGTCTAAATGACTCCTTGATCCATAAGCTGCAGAATGGGTGTTGTG 2303


RESULT 21
AK095077
LOCUS       AK095077               3895 bp    mRNA    linear   PRI 30-JAN-2004
DEFINITION  Homo sapiens cDNA FLJ37758 fis, clone BRHIP2023869.
ACCESSION   AK095077
VERSION     AK095077.1  GI:21754267
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
            Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
            Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
            Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
            Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
            Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
            Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
            Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
            Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
            Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
            Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
            Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,
            Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
            Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
            Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
            Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
            Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
            Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
            Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
            Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
            Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
            Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
```

```
            Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
            Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
            Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
            Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
            Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
  TITLE     Complete sequencing and characterization of 21,243 full-length
            human cDNAs
  JOURNAL   Nat. Genet. 36 (1), 40-45 (2004)
   PUBMED   14702039
REFERENCE   2
  AUTHORS   Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
            Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
            Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3895)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1. .3895
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="BRHIP2023869"
                     /tissue_type="hippocampus"
                     /clone_lib="BRHIP2"
                     /note="cloning vector: pME18SFL3"
ORIGIN

  Query Match             25.8%;  Score 516;  DB 8;  Length 3895;
  Best Local Similarity   77.5%;  Pred. No. 1e-70;
  Matches  741;  Conservative    0;  Mismatches  175;  Indels   40;  Gaps    8;

Qy       1071 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 1130
             | || | |||||| |||||||| ||| |||||||| |  ||| |||| ||| |||||| |
Db       1362 ATAAATTTTTTGGTTTCCCAGTACATGTAAAAGTTATAATTACACTATACTATAGTCTAT 1421

Qy       1131 TAAGTGTGCAATAGTGTTATGTCTAAAAAAA------CACATACCTTAATTTTAAAATGC 1184
              |||| |||||||||  |||||||| ||||||        ||| |||||||| ||||| |
Db       1422 TAAGGGTGCAATAGCATTATGTCTGAAAAAAACAATGGTCATGCCTTAATTTAAAAATAC 1481

Qy       1185 TTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCT 1244
              |||||||||||||||||| ||||||| ||| | |||||||||  |||| ||||||||
Db       1482 TTTATTACTAAAAAATGCTAGCAATCATCTGAACCTTCAGTGAGTCATAATATTTTTGCT 1541

Qy       1245 GGTGGAAGGTCTTTTCTTATTGATGACTGAT----------CGGGGGTCAGGTGCTGAA 1293
               ||||| || |||  ||   || |||  | |           ||  |   | |||||||
Db       1542 AGTGGAGGGCCTTGCCTGGATGTTGATGGCTGCTGACTGATCAGGATGGTGGTTGCTGAA 1601

Qy       1294 GCTTAGGGTGGCTGTGGCAGTTTCT-----TAAAACAACAGTGAAGATTGCAATATCAGT 1348
              | |||||||||||| |||| |||||     ||| ||||||||||| || |   ||    |
Db       1602 GGTTAGGGTGGCTGCGGCAATTTCTTAACATAAGACAACAGTGAAGTTTCCTGCATTGAT 1661

Qy       1349 TGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTA 1408
              |||||||||||||||||||| |||||||| |||   ||| ||||| ||||||||||||||
Db       1662 TGACTCTTCCTTTCATGAAATATTTCTCTATAGCATGTAATGCTGTTTGATAGCATTTTA 1721

Qy       1409 TGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAA 1467
              |||||||||| || ||| | ||||| | ||| |||||| ||||||||       || |
Db       1722 CCCACAGTAGAATTTATTTCAGAATTGAAGTCAGTCCTCTTAAACCCTGTCACTTCTTTA 1781

Qy       1468 CAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTG 1527
              | ||||||| |   |||||||| ||||| ||||||||||||||||||| ||||||
Db       1782 TTAACTAAGTTCACGCAATATTCTAAATCCTTTGTTGTCATTTCAACAATGTTCACAACA 1841

Qy       1528 TCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAG 1587
              |||||||||||  |||||||  |||||        | |    |||||| |||| | ||||
Db       1842 TCTTCACCAGGGATAGATTCTGTCTCAAGAAACCACTT----TCTTTGTTCATTCGTAAG 1897

Qy       1588 AAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTC 1647
              ||| ||  |||||||||||||||||| |||||||||| |||||| ||||||  | |||
Db       1898 AAGCAACCCCTCATCTGTTCAAGTTTTCTCATGAGATTGGAGCAATTCAGTCACATTTTC 1957

Qy       1648 AGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCT 1707
              |||      || || |||||||||| |||||||||||| ||||||||||||||  ||| ||
Db       1958 AGG-----CTGCATTTTTAATTCTAGTTCTCTTGCTATTTCTACCACATCTGCAGTTACT 2012

Qy       1708 TCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTC 1767
              |||| || ||||| |||||||| ||| |||||| |||||||||||||||| ||||||||
Db       2013 TCCTTCACTGAAGGCTTGAACCCCTCAAAGTCACCCATGAGGGTTGGAATCAACTTCTTC 2072

Qy       1768 CAAATTCCTGTTAATATTTATATTTTGACCTCC---CATGAATCATGAATGTTCTTAATG 1824
              |||||||||||||||| || || ||||||||||    ||||||| |||||||| ||||||
Db       2073 CAAATTCCTGTTAATGTTGATGTTTTGACCTCCTCTCATGAATGATGAATGTTTTTAATG 2132

Qy       1825 GCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCC 1884
              || ||  |||||| | |||||||||| |||||||||||||||||| | ||  ||||||
Db       2133 ACATCTAAAATGGTCAGTCCTTTCCAGAAGGTTTTCAATTTACTTTGCCCCAATCCAT-- 2190

Qy       1885 ATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAAT 1944
                |||||||  |||| || ||| ||  | ||| |||||  || |||||||| ||||||
Db       2191 ---CAGAGGAATCACTGTCTATGGCAACTGTAGGCTTATAAAAGGTATTTCTTAAATAAT 2247

Qy       1945 AAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              | | |||||||||  |||| ||||||||||||||   ||||| ||| | |||||||
Db       2248 ATGTCTTGAAAGTCTAAATGACTCCTTGATCCATAAGCTGCAGAATGGGTGTTGTG 2303


RESULT 22
AC004736/c
LOCUS       AC004736               115958 bp    DNA     linear   PRI 03-JUL-1998
DEFINITION  Human Chromosome 11p14.3 PAC clone pDJ1082L12 containing KNCN1 and
            MyoD, complete sequence.
ACCESSION   AC004736
VERSION     AC004736.1  GI:3282183
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 115958)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 115958)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,C., Dunn,J.,
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
            McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-1998) Genome Science & Technology Center,
```

```
            University of Texas Southwestern Medical Center, 5323 Harry Hines
            Blvd, Dallas, TX 75235-8591, USA
REFERENCE   3  (bases 1 to 115958)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1998) Genome Science & Technology Center,
            University of Texas Southwestern Medical Center, 5323 Harry Hines
            Blvd, Dallas, TX 75235-8591, USA
REFERENCE   4  (bases 1 to 115958)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1998) Genome Science & Technology Center,
            University of Texas Southwestern Medical Center, 5323 Harry Hines
            Blvd, Dallas, TX 75235-8591, USA
COMMENT     On Jul 1, 1998 this sequence version replaced gi:3152625.
            NOTE: Extent of pDJ1082L12 overlap with mapped clones; SP6 PAC
            cloning end: 1. .8775, pDJ59m18; T7 PAC cloning end: 45252. .115958,
            6-109h6.  Further information regarding the map of this region or
            annotation of pDJ1082L12 can be found at
            http://gestec.swmed.edu/chromoso2.htm.
            CHROMSOMAL LOCUS: This PAC clone comes from the Usher syndrome
            region (USH1C) mapped between STSs markers D11S1310 and 1115A14.
            MARKER CONFIRMATION: STSs/ESTs sequence confirmed; D11S4130, KNCN1,
            and MyoD1.
            MAPPED CLONE OVERLAP: pDJ59m18 and 6-109h6.
            IMPORTANT: This submission contains the entire insert of clone
            pDJ1082L12.  pDJ1082L12 comes from a PAC library constructed atthe
            Roswell Park Cancer Institute by the Pieter de Jong group.  This
            clone has been finished according to strict quality criteria and
            attempts have been made to resolve all base calling problems such
            as compressions and repetitive elements. The expected Phred/Phrap
            calculated errors/10kb is 0.39.  In addition, this sequence has
            been finished such that 99.9% of consensus base calls consist of
            either double-stranded coverage or 2 types of labeling chemistry on
            one strand.
FEATURES             Location/Qualifiers
     source          1. .115958
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     repeat_region   4409. .4517
                     /rpt_family="MIR"
     repeat_region   complement(19171. .19291)
                     /rpt_family="Alu"
     repeat_region   complement(21561. .21985)
                     /rpt_family="Alu"
     repeat_region   23678. .23767
                     /rpt_family="MIR"
     repeat_region   28877. .31250
                     /rpt_family="Tigger1"
     repeat_region   complement(31777. .31876)
                     /rpt_family="MIR"
     repeat_region   complement(36023. .36231)
                     /rpt_family="MIR"
     repeat_region   complement(39712. .39766)
                     /rpt_family="MIR"
     repeat_region   complement(41971. .42073)
                     /rpt_family="MIR"
     repeat_region   45642. .45752
                     /rpt_family="MIR"
     repeat_region   57204. .57337
                     /rpt_family="MIR"
     repeat_region   complement(58027. .58249)
                     /rpt_family="MER20"
     repeat_region   59316. .59729
                     /rpt_family="L1"
     repeat_region   complement(60013. .60386)
                     /rpt_family="THE1"
     repeat_region   60398. .60596
                     /rpt_family="L1"
     repeat_region   60940. .61228
                     /rpt_family="Alu"
     repeat_region   complement(63635. .63900)
                     /rpt_family="Alu"
     repeat_region   complement(75510. .75774)
                     /rpt_family="Alu"
     repeat_region   complement(82106. .82378)
                     /rpt_family="Alu"
     repeat_region   complement(84150. .84211)
                     /rpt_family="MLT1"
     repeat_region   84727. .85172
                     /rpt_family="MLT1"
     repeat_region   85852. .86134
                     /rpt_family="Alu"
     repeat_region   86387. .86565
                     /rpt_family="Alu"
     repeat_region   complement(88437. .88514)
                     /rpt_family="MIR"
     repeat_region   complement(90686. .90752)
                     /rpt_family="MIR"
     repeat_region   91042. .91163
                     /rpt_family="MIR"
     repeat_region   94274. .94419
                     /rpt_family="MIR"
     repeat_region   complement(95470. .95617)
                     /rpt_family="MLT1"
     repeat_region   complement(97484. .97917)
                     /rpt_family="L1"
     repeat_region   105586. .105657
                     /rpt_family="MIR"
     repeat_region   105673. .105884
                     /rpt_family="MIR"
     repeat_region   complement(106369. .106641)
                     /rpt_family="Alu"
     repeat_region   complement(107302. .107596)
                     /rpt_family="Alu"
     repeat_region   109188. .109458
                     /rpt_family="Alu"
     repeat_region   complement(110473. .110773)
                     /rpt_family="Alu"
     repeat_region   complement(111227. .111337)
                     /rpt_family="L1"
     repeat_region   111412. .111792
                     /rpt_family="THE1"
     repeat_region   complement(112629. .112992)
                     /rpt_family="L1"
ORIGIN

  Query Match            25.8%;  Score 515.2;  DB 8;  Length 115958;
  Best Local Similarity  75.6%;  Pred. No. 3.9e-71;
  Matches  789;  Conservative    0;  Mismatches  203;  Indels   52;  Gaps   10;

Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
              ||||||||||||| || |||| |||||| ||||||||| |||||  |||||
Db      31252 TACAGGCATACCTCGGTGATATTGTGGGCTTGGTTCCAGACCACAGCAATA--------- 31202

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096
                ||||| ||||||||||||||  ||||||| |  | |||||| ||| ||| ||||
Db      31201 ------AAGTGAATATCACAATAAAGCAAGTCACATA--TTTTTTGGTTTCTCAGAGCAT 31150

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156
              |||||||| |  |||||||| ||||||   | |||| ||||||||||  ||||||||||
Db      31149 ATAAAAGTTATATTTATACTATACTGTACATTATTAATTGTGCAATAGCATTATGTCTAA 31090

Qy       1157 AAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGA 1216
              |||   ||||| ||||||||| |||| | ||||| ||  |||||||||||||||||| |||
```

```
Db      31089 AAATGTACATATCTTAATTTTTAAATACGTTATTGCTTCAAAATGCTAACAATCATATGA 31030

Qy       1217 GCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT- 1275
              || ||||||||||  |||||||||||||| | || ||||||  || | || |||  | |
Db      31029 GCCTTCAGTGAGTCATAATCTTTTTGCTGATAGAGGGTCTTGCCTCAGTGTTGATGGCTG 30970

Qy       1276 ----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA--- 1322
                        | | |      |||||||| || ||||||    | ||| |||  ||||
Db      30969 CCGATTGATCAGCGTGGTGATTGCTGAAGGTTGGGGTGGACATAGCAATTTTGTAAAGTA 30910

Qy       1323 --ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
                ||||||  ||||  ||||    ||||  |||||||| |||||||  |||||||||||| ||
Db      30909 AGACAACAAAGAAGTTTGCTGCATCAATTGACTCTCCCTTTCACAAAAGATTTCTCTGTA 30850

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TC 1439
              |  |||||||||  |||||  ||||||  || ||||||||||||||||  |||| || | ||
Db      30849 GCATGTGATGCTGTTTGACAGCATTCCATCCACAGTAGAACTTCTTTCAAAACTGAAGTC 30790

Qy       1440 AATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCAT 1499
              ||||||||||||||||||     || |  | || |||| ||  |||||||| ||||| |
Db      30789 AATCCTCTCAAACCCTGCTGCTGCTTTATCAGCTGAGTTTATGGAATATTCTAAATCCTT 30730

Qy       1500 TGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCC 1559
              |||||||||||||||||  || | ||   |||||||||||| ||||||||||||||
Db      30729 TGTTGTCATTTCAACAATGTTTATAGCATCTTCACCAGGAATAGATTCCATCTCAAGAAA 30670

Qy       1560 TGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCAT 1619
                |  |    |||||||| |||||||||||| || ||||||||||||||||||| |||||
Db      30669 CCACTT----TCTTTGCTTATCCATAAGAAGCAACTCCTCATCTGTTCAAGTTTGATCAT 30614

Qy       1620 GAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCT 1679
              |||||||||||||| ||||||  ||||||||      ||  |||| ||||||  || |||||
Db      30613 GAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCTACTTCTAATTCTAGCTCTCT 30559

Qy       1680 TGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTC 1739
              |||| |||  | ||||||||  ||  ||||||||| || ||||||||||  |  |||||
Db      30558 TGCTATTTTCATCACATCTGCAGTGACTTCCTCCACTGGAGTCTTGAACCCATTAAAGTC 30499

Qy       1740 ATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---C 1796
              || || |||  |||||||| ||| |||   || ||||||||| ||  | ||||||    |
Db      30498 ATTCACGAGAATTGGAATCAACTCCTTTTGAACTCCTGTTAATGTTGCTGTTTTGACCTC 30439

Qy       1797 CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGT 1856
              |||| ||||||||  | |||||||| || || || ||||||| ||| ||||||| ||||
Db      30438 CTCCGATGAATCACAAGTGTTCTTAGTGACATCTAGAATGGTAAATTCTTTCCAGAAGG- 30380

Qy       1857 TTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATA 1916
              ||||||| |||| |  |||||  ||     |||||||  || | || |||  || || |
Db      30379 TTTCAATTGACTTTGCACAGATTGAT-----CAGAGGAATCAGTATCTATGGAAGCTACA 30325

Qy       1917 GCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCC 1976
              |||||||| |||||||||||| || |||| | |||||||||  |||| |||||||| |||
Db      30324 GCCTTATGAAATGTATTTCTTAAAGAATATGACTTGAAAGTAAAAATGACTCCTTGGTCC 30265

Qy       1977 ATTTTCTGCAAAATAGATGTTGTG 2000
              ||   || || ||| | || || |
Db      30264 ATGGGCTTCAGAATGGGTGCTGAG 30241
```

```
RESULT 23
AC124056
LOCUS       AC124056               115971 bp    DNA     linear   PRI 21-DEC-2002
DEFINITION  Homo sapiens chromosome 11, clone RP5-1082L12, complete sequence.
ACCESSION   AC124056
VERSION     AC124056.8  GI:27356724
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 115971)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP5-1082L12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 115971)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 115971)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 115971)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 21, 2002 this sequence version replaced gi:25102226.
```

```
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L27400
              Center clone name: 1082_L_12
          --------------.
FEATURES             Location/Qualifiers
     source          1. .115971
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP5-1082L12"
                     /clone_lib="RPCI Human PAC library 5"
     repeat_region   99. .1108
                     /rpt_family="L2"
     repeat_region   1206. .1473
                     /rpt_family="L2"
     repeat_region   complement(1602. .1704)
                     /rpt_family="L1ME"
     repeat_region   complement(1845. .2046)
                     /rpt_family="MIR3"
     repeat_region   complement(2352. .2469)
                     /rpt_family="MIR"
     repeat_region   2612. .4114
                     /rpt_family="L1MC1"
     repeat_region   4083. .4170
                     /rpt_family="L1ME3A"
     repeat_region   complement(4171. .4551)
                     /rpt_family="MSTA"
     repeat_region   4552. .4920
                     /rpt_family="L1ME3A"
     repeat_region   4925. .4986
                     /rpt_family="L1ME3A"
     repeat_region   4970. .5179
                     /rpt_family="L1MD3"
     repeat_region   5180. .5490
                     /rpt_family="AluSx"
     repeat_region   5491. .5609
                     /rpt_family="L1MD3"
     repeat_region   5609. .5845
                     /rpt_family="L1ME3A"
     repeat_region   complement(6442. .6757)
                     /rpt_family="AluSx"
     repeat_region   8356. .8671
                     /rpt_family="AluSx"
     unsure          8717. .8725
                     /note="<30 qual SNGL region"
     unsure          8735. .8770
                     /note="<30 qual SNGL region"
     unsure          8786. .8790
                     /note="<30 qual SNGL region"
     repeat_region   9317. .9609
                     /rpt_family="AluSc"
     repeat_region   complement(10070. .10405)
                     /rpt_family="MIR"
     repeat_region   complement(11554. .11859)
                     /rpt_family="LTR16C"
     repeat_region   12223. .12374
                     /rpt_family="MIR"
     repeat_region   13250. .13353
                     /rpt_family="MIR3"
     repeat_region   complement(14122. .14358)
                     /rpt_family="MIR"
     repeat_region   14653. .14758
                     /rpt_family="MIR"
     repeat_region   16324. .17416
                     /rpt_family="L2"
     repeat_region   complement(17963. .18020)
                     /rpt_family="MLT1E2"
     repeat_region   18027. .18705
                     /rpt_family="L1MB5"
     repeat_region   complement(18706. .19168)
                     /rpt_family="MLT1E2"
     repeat_region   19383. .19447
                     /rpt_family="L2"
     repeat_region   complement(20015. .20094)
                     /rpt_family="L2"
     repeat_region   20108. .20563
                     /rpt_family="MLT1F"
     repeat_region   21080. .21223
                     /rpt_family="L3"
     repeat_region   complement(21504. .21725)
                     /rpt_family="MIR"
     repeat_region   complement(22444. .22513)
                     /rpt_family="MIR3"
     repeat_region   22970. .23148
                     /rpt_family="G-rich"
     repeat_region   24021. .24120
                     /rpt_family="L2"
     repeat_region   complement(24730. .24934)
                     /rpt_family="MIR"
     repeat_region   24947. .25068
                     /rpt_family="L2"
     repeat_region   25137. .25277
                     /rpt_family="MIR"
     repeat_region   25441. .25539
                     /rpt_family="MIR"
     repeat_region   26018. .26087
                     /rpt_family="L2"
     repeat_region   27434. .27568
                     /rpt_family="MIR"
     repeat_region   complement(28003. .28114)
                     /rpt_family="MIR"
     repeat_region   28115. .28608
                     /rpt_family="MLT1H"
     repeat_region   complement(28622. .29055)
                     /rpt_family="L1MC1"

  Query Match             25.8%;  Score 515.2;  DB 8;  Length 115971;
  Best Local Similarity   75.6%;  Pred. No. 3.9e-71;
  Matches  789;  Conservative    0;  Mismatches  203;  Indels   52;  Gaps   10;

Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
              |||||||||||||  || ||||  ||||||  |||||||| |||||  |||||
Db      84711 TACAGGCATACCTCGGTGATATTGTGGGCTTGGTTCCAGACCACAGCAATA--------- 84761

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096
                |||||  ||||||||||||||  |||||||  |  | ||||||  ||| ||| ||||
Db      84762 ------AAGTGAATATCACAATAAAGCAAGTCACATA--TTTTTTGGTTTCTCAGAGCAT 84813

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156
              |||||||||  |  ||||||||  ||||||    | ||||  ||||||||||  ||||||||||
Db      84814 ATAAAAGTTATATTTATACTATACTGTACATTATTAATTGTGCAATAGCATTATGTCTAA 84873

Qy       1157 AAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGA 1216
              |||   |||||  |||||||||  ||||  |  |||||  ||  ||||||||||||||||  |||
Db      84874 AAATGTACATATCTTAATTTTTAAATACGTTATTGCTTCAAAATGCTAACAATCATATGA 84933

Qy       1217 GCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT- 1275
              ||  ||||||||||  ||||||||||||||  |  ||  ||||||  ||  |  ||  |||  |  |
Db      84934 GCCTTCAGTGAGTCATAATCTTTTTGCTGATAGAGGGTCTTGCCTCAGTGTTGATGGCTG 84993

Qy       1276 ----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA--- 1322
                       | | |    ||||||||  ||  ||||||    |  |||  |||  ||||
Db      84994 CCGATTGATCAGCGTGGTGATTGCTGAAGGTTGGGGTGGACATAGCAATTTTGTAAAGTA 85053

Qy       1323 --ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
                ||||||  ||||  ||||   ||||  ||||||||  |||||||  ||||||||||||  ||
```

```
Db      85054 AGACAACAAAGAAGTTTGCTGCATCAATTGACTCTCCCTTTCACAAAAGATTTCTCTGTA 85113

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TC 1439

Db      85114 GCATGTGATGCTGTTTGACAGCATTCCATCCACAGTAGAACTTCTTTCAAAACTGAAGTC 85173

Qy       1440 AATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCAT 1499

Db      85174 AATCCTCTCAAACCCTGCTGCTGCTTTATCAGCTGAGTTTATGGAATATTCTAAATCCTT 85233

Qy       1500 TGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCC 1559

Db      85234 TGTTGTCATTTCAACAATGTTTATAGCATCTTCACCAGGAATAGATTCCATCTCAAGAAA 85293

Qy       1560 TGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCAT 1619

Db      85294 CCACTT----TCTTTGCTTATCCATAAGAAGCAACTCCTCATCTGTTCAAGTTTGATCAT 85349

Qy       1620 GAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCT 1679

Db      85350 GAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCTACTTCTAATTCTAGCTCTCT 85404

Qy       1680 TGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTC 1739

Db      85405 TGCTATTTTCATCACATCTGCAGTGACTTCCTCCACTGGAGTCTTGAACCCATTAAAGTC 85464

Qy       1740 ATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---C 1796

Db      85465 ATTCACGAGAATTGGAATCAACTCCTTTTGAACTCCTGTTAATGTTGCTGTTTTGACCTC 85524

Qy       1797 CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGT 1856

Db      85525 CTCCGATGAATCACAAGTGTTCTTAGTGACATCTAGAATGGTAAATTCTTTCCAGAAGG- 85583

Qy       1857 TTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATA 1916

Db      85584 TTTCAATTGACTTTGCACAGATTGAT-----CAGAGGAATCAGTATCTATGGAAGCTACA 85638

Qy       1917 GCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCC 1976

Db      85639 GCCTTATGAAATGTATTTCTTAAAGAATATGACTTGAAAGTAAAAATGACTCCTTGGTCC 85698

Qy       1977 ATTTTCTGCAAAATAGATGTTGTG 2000

Db      85699 ATGGGCTTCAGAATGGGTGCTGAG 85722
```

RESULT 24
AC124301/c
LOCUS AC124301 145713 bp DNA linear PRI 30-AUG-2002
DEFINITION Homo sapiens chromosome 11, clone RP11-358H18, complete sequence.
ACCESSION AC124301
VERSION AC124301.6 GI:22549794
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145713)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-358H18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145713)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 145713)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 145713)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Aug 30, 2002 this sequence version replaced gi:22507191.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L27330
Center clone name: 358_H_18
--------------.

```
FEATURES             Location/Qualifiers
     source          1. .145713
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-358H18"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   539. .652
                     /rpt_family="MIR3"
     repeat_region   complement(1287. .1371)
                     /rpt_family="MER5A"
     repeat_region   1477. .1595
                     /rpt_family="MIR"
     repeat_region   complement(2216. .2346)
                     /rpt_family="MIR"
     repeat_region   3134. .5515
                     /rpt_family="Tigger1"
     repeat_region   complement(5535. .5850)
                     /rpt_family="L3"
     repeat_region   complement(6037. .6228)
                     /rpt_family="MIR"
     repeat_region   6565. .6605
                     /rpt_family="(TG)n"
     repeat_region   complement(6687. .7013)
                     /rpt_family="L2"
     repeat_region   7025. .8159
                     /rpt_family="L2"
     unsure          complement(7232. .7259)
                     /note="single clone coverage"
     unsure          complement(7313. .7807)
                     /note="single clone coverage"
     unsure          complement(7313. .7392)
                     /note="<30 qual single clone coverage"
     unsure          complement(7417. .7453)
                     /note="<30 qual single clone coverage"
     unsure          complement(7441. .7445)
                     /note="probably CCCCC; possibly CCCCCC"
     unsure          complement(7480. .7485)
                     /note="<30 qual single clone coverage"
     unsure          complement(7506. .7512)
                     /note="<30 qual single clone coverage"
     unsure          7568. .7572
                     /note="<30 qual single clone coverage"
     unsure          7931. .8083
                     /note="single clone coverage"
     unsure          7945. .7960
                     /note="<30 qual single clone coverage"
     unsure          8049. .8078
                     /note="<30 qual single clone coverage"
     unsure          8087. .8098
                     /note="<30 qual SNGL region"
     repeat_region   complement(8200. .8285)
                     /rpt_family="MIR"
     repeat_region   8318. .8400
                     /rpt_family="L2"
     repeat_region   9863. .9995
                     /rpt_family="MIR3"
     repeat_region   complement(10135. .10185)
                     /rpt_family="L2"
     repeat_region   complement(10262. .10501)
                     /rpt_family="MIR"
     repeat_region   11692. .11851
                     /rpt_family="(TCCA)n"
     repeat_region   complement(11884. .11978)
                     /rpt_family="MIR3"
     repeat_region   11962. .12097
                     /rpt_family="MIR"
     repeat_region   complement(12299. .12507)
                     /rpt_family="L2"
     repeat_region   complement(12896. .13116)
                     /rpt_family="MIR"
     repeat_region   complement(13889. .14059)
                     /rpt_family="MIR"
     repeat_region   complement(14500. .14724)
                     /rpt_family="L1PB3"
     repeat_region   14994. .15128
                     /rpt_family="L2"
     repeat_region   complement(15845. .15940)
                     /rpt_family="MIR"
     repeat_region   complement(16127. .16338)
                     /rpt_family="MIR"
     repeat_region   16552. .16622
                     /rpt_family="G-rich"
     repeat_region   complement(17059. .17839)
                     /rpt_family="L2"
     repeat_region   complement(17984. .18134)
                     /rpt_family="MER5A"
     repeat_region   complement(18889. .19018)
                     /rpt_family="MIR3"
     repeat_region   19824. .20072
                     /rpt_family="MIR"
     repeat_region   20312. .20429
                     /rpt_family="MIR3"
     repeat_region   20927. .21071
                     /rpt_family="MIR"
     repeat_region   complement(21199. .21657)
                     /rpt_family="L2"
     repeat_region   complement(21709. .21849)
                     /rpt_family="MIR"
     repeat_region   22459. .22485
                     /rpt_family="(CA)n"
```

```
  Query Match            25.8%;  Score 515.2;  DB 8;  Length 145713;
  Best Local Similarity  75.6%;  Pred. No. 3.6e-71;
  Matches  789;  Conservative    0;  Mismatches  203;  Indels   52;  Gaps   10;

Qy       977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
             ||||||||||||| || |||| |||||| |||||||| |||||  |||||
Db      5517 TACAGGCATACCTCGGTGATATTGTGGGCTTGGTTCCAGACCACAGCAATA--------- 5467

Qy      1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096
                |||| ||||||||||||  |||||| |  | |||||| ||| ||| ||||
Db      5466 ------AAGTGAATATCACAATAAAGCAAGTCACATA--TTTTTTGGTTTCTCAGAGCAT 5415

Qy      1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156
             ||||||||| |  |||||||| ||||||   | |||| ||||||||||  ||||||||||
Db      5414 ATAAAAGTTATATTTATACTATACTGTACATTATTAATTGTGCAATAGCATTATGTCTAA 5355

Qy      1157 AAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGA 1216
             |||   ||||| ||||||||| |||| | ||||| ||  |||||||||||||||||| |||
Db      5354 AAATGTACATATCTTAATTTTTAAATACGTTATTGCTTCAAAATGCTAACAATCATATGA 5295

Qy      1217 GCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT- 1275
             || |||||||||  |||||||||||||| | || ||||||  || | || |||  | |
Db      5294 GCCTTCAGTGAGTCATAATCTTTTTGCTGATAGAGGGTCTTGCCTCAGTGTTGATGGCTG 5235

Qy      1276 ----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA--- 1322
                       | | |     |||||||| || ||||||   | ||| |||  ||||
Db      5234 CCGATTGATCAGCGTGGTGATTGCTGAAGGTTGGGGTGGACATAGCAATTTTGTAAAGTA 5175

Qy      1323 --ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
               ||||||  |||| ||||   |||| |||||||| |||||||  ||||||||||||| ||
Db      5174 AGACAACAAAGAAGTTTGCTGCATCAATTGACTCTCCCTTTCACAAAAGATTTCTCTGTA 5115

Qy      1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TC 1439
             |  ||||||||| ||||| ||||||  || ||||||||||||||||| |||| || | ||
Db      5114 GCATGTGATGCTGTTTGACAGCATTCCATCCACAGTAGAACTTCTTTCAAAACTGAAGTC 5055

Qy      1440 AATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCAT 1499
             ||||||||||||||||||     || |  | || |||| ||  |||||||| ||||| |
Db      5054 AATCCTCTCAAACCCTGCTGCTGCTTTATCAGCTGAGTTTATGGAATATTCTAAATCCTT 4995

Qy      1500 TGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCC 1559
             |||||||||||||||||| || | ||  |||||||||||| ||||||||||||||
```

```
Db       4994 TGTTGTCATTTCAACAATGTTTATAGCATCTTCACCAGGAATAGATTCCATCTCAAGAAA 4935

Qy       1560 TGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCAT 1619

Db       4934 CCACTT----TCTTTGCTTATCCATAAGAAGCAACTCCTCATCTGTTCAAGTTTGATCAT 4879

Qy       1620 GAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCT 1679

Db       4878 GAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCTACTTCTAATTCTAGCTCTCT 4824

Qy       1680 TGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTC 1739

Db       4823 TGCTATTTTCATCACATCTGCAGTGACTTCCTCCACTGGAGTCTTGAACCCATTAAAGTC 4764

Qy       1740 ATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---C 1796

Db       4763 ATTCACGAGAATTGGAATCAACTCCTTTTGAACTCCTGTTAATGTTGCTGTTTTGACCTC 4704

Qy       1797 CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGT 1856

Db       4703 CTCCGATGAATCACAAGTGTTCTTAGTGACATCTAGAATGGTAAATTCTTTCCAGAAGG- 4645

Qy       1857 TTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATA 1916

Db       4644 TTTCAATTGACTTTGCACAGATTGAT-----CAGAGGAATCAGTATCTATGGAAGCTACA 4590

Qy       1917 GCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCC 1976

Db       4589 GCCTTATGAAATGTATTTCTTAAAGAATATGACTTGAAAGTAAAAATGACTCCTTGGTCC 4530

Qy       1977 ATTTTCTGCAAAATAGATGTTGTG 2000

Db       4529 ATGGGCTTCAGAATGGGTGCTGAG 4506
```

RESULT 25
HS168L15/c
LOCUS HS168L15 178098 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP1-168L15 on chromosome 6q26-27 Contains the 3' end of the RPS6KA2 gene for ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3), the gene for brain protein 44-like (BRP441), the 5' end of the gene for a novel protein similar to RIKEN cDNA 5630401J11 (pRGR1) (MGC19825) and 5 CpG islands, complete sequence.
ACCESSION AL022069
VERSION AL022069.1 GI:3256174
KEYWORDS HTG; BRP441; CpG island; MGC19825; pRGR1; RPS6KA2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 178098)
AUTHORS Williams,S.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 26, 1998 this sequence version replaced gi:2950404.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6
RP1-168L15 is from the library RPCI-1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vega@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES Location/Qualifiers

source 1. .178098
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIP704L15168"
/db_xref="taxon:9606"
/chromosome="6"
/map="q26-27"
/clone="RP1-168L15"
/clone_lib="RPCI-1"

gene join(Z98049.1:25497. .25817,Z98049.1:113954. .114070, Z98049.1:121425. .121506,6546. .6626,8621. .8700, 12291. .12397,15979. .16016,18253. .18395,26162. .26232, 27982. .28070,46998. .47062,57352. .57454,65670. .65800, 68054. .68179,84413. .84502,86292. .86450,93486. .93647, 96945. .97021,98561. .98678,102972. .103109,104016. .107517)
/gene="RPS6KA2"
/locus_tag="RP1-168L15.2-001"

mRNA join(Z98049.1:25497. .25817,Z98049.1:113954. .114070, Z98049.1:121425. .121506,6546. .6626,8621. .8700, 12291. .12397,15979. .16016,18253. .18395,26162. .26232, 27982. .28070,46998. .47062,57352. .57454,65670. .65800, 68054. .68179,84413. .84502,86292. .86450,93486. .93647, 96945. .97021,98561. .98678,102972. .103109,104016. .107517)
/gene="RPS6KA2"
/locus_tag="RP1-168L15.2-001"
/product="ribosomal protein S6 kinase, 90kDa, polypeptide 2"
/note="match: ESTs: AA409581 AA447506 AI903982 AJ453645 AL041466 AL517867 AL523835 AL523836 AU117630 AU130117 AU144411 AW410244 AW503268 AW654413 BE653200 BF205134 BF364359 BF513415 BF982517 BG295860 BG716876 BG949099 BI084095 BI198735 BI335924 BI836819 BM669400 BM699585 BM750954 BM766849 BM801851 BM873603 BQ029058 BQ180508 BQ277933 BQ303855 BQ421428 BQ430681 BQ431157 F06080 N79949 T74714 Z45429
match: cDNAs: AF140710 AJ131021 AK019881 BC002363 BC011189 BC014966 D83013 L07597 L07598 X85106"

CDS join(Z98049.1:25719. .25817,Z98049.1:113954. .114070, Z98049.1:121425. .121506,6546. .6626,8621. .8700, 12291. .12397,15979. .16016,18253. .18395,26162. .26232, 27982. .28070,46998. .47062,57352. .57454,65670. .65800, 68054. .68179,84413. .84502,86292. .86450,93486. .93647, 96945. .97021,98561. .98678,102972. .103109,104016. .>104140)
/gene="RPS6KA2"
/locus_tag="RP1-168L15.2-001"
/standard_name="OTTHUMP00000017601"
/note="match: proteins: BAB85907 P18653 P18654 P51812 Q15349 Q21734 Q63531 Q96J38 Q9BE29 Q9D2C0 Q9PTN1 Q9WUT3"
/codon_start=1
/product="ribosomal protein S6 kinase, 90kDa, polypeptide 2"
/protein_id="CAI19651.1"
/db_xref="GI:56202585"
/db_xref="GOA:Q15349"
/db_xref="UniProt/Swiss-Prot:Q15349"
/translation="MDLSMKKFAVRRFFSVYLRRKSRSKSSSLSRLEEEGVVKEIDIS HHVKEGFEKADPSQFELLKVLGQGSYGKVFLVRKVKGSDAGQLYAMKVLKKATLKVRD RVRSKMERDILAEVNHPFIVKLHYAFQTEGKLYLILDFLRGGDLFTRLSKEVMFTEED VKFYLAELALALDHLHSLGIIYRDLKPENILLDEEGHIKITDFGLSKEAIDHDKRAYS FCGTIEYMAPEVVNRRGHTQSADWWSFGVLMFEMLTGSLPFQGKDRKETMALILKAKL

```
                GMPQFLSGEAQSLLRALFKRNPCNRLGAGIDGVEEIKRHPFFVTIDWNTLYRKEIKPP
                FKPAVGRPEDTFHFDPEFTARTPTDSPGVPPSANAHHLFRGFSFVASSLIQEPSQQDL
                HKVPVHPIVQQLHGNNIHFTDGYEIKEDIGVGSYSVCKRCVHKATDTEYAVKIIDKSK
                RDPSEEIEILLRYGQHPNIITLKDVYDDGKFVYLVMELMRGGELLDRILRQRYFSERE
                ASDVLCTITKTMDYLHSQGVVHRDLKPSNILYRDESGSPESIRVCDFGFAKQLRAGNG
                LLMTPCYTANFVAPEVLKRQGYDAACDIWSLGILLYTMLAGFTPFANGPDDTPEEILA
                RIGSGKYALSGGNWDSISDAAKDVVSKMLHVDPHQRLTAMQVLKHPWVVNREYLSPNQ
                LSRQDVHLVKGAMAATYFALNRTPQAPRLEPVLSSNLAQRRGMKRLTSTRL"
gene            join(6593. .6626,8621. .8700,12291. .12397,15979. .16016,
                18253. .18395,26162. .26232,27982. .28070,41064. .41177)
                /gene="RPS6KA2"
                /locus_tag="RP1-168L15.2-002"
mRNA            join(6593. .6626,8621. .8700,12291. .12397,15979. .16016,
                18253. .18395,26162. .26232,27982. .28070,41064. .41177)
                /gene="RPS6KA2"
                /locus_tag="RP1-168L15.2-002"
                /product="ribosomal protein S6 kinase, 90kDa, polypeptide
                2"
                /note="match: ESTs: BE857569"
CDS             join(12294. .12397,15979. .16016,18253. .18395,26162. .26232,
                27982. .28070,41064. .41119)
                /gene="RPS6KA2"
                /locus_tag="RP1-168L15.2-002"
                /standard_name="OTTHUMP00000017602"
                /codon_start=1
                /product="ribosomal protein S6 kinase, 90kDa, polypeptide
                2"
                /protein_id="CAI19650.1"
                /db_xref="GI:56202584"
                /db_xref="GOA:Q5TI62"
                /db_xref="InterPro:IPR000719"
                /db_xref="InterPro:IPR001245"
                /db_xref="InterPro:IPR002290"
                /db_xref="InterPro:IPR008271"
                /db_xref="UniProt/TrEMBL:Q5TI62"
                /translation="MFTEEDVKFYLAELALALDHLHSLGIIYRDLKPENILLDEEGHI
                KITDFGLSKEAIDHDKRAYSFCGTIEYMAPEVVNRRGHTQSADWWSFGVLMFEMLTGS
                LPFQGKDRKETMALILKAKLGMPQFLSGEAQSLLRALFKRNPCNRLGEKTALCPRVSC
                SRLLGF"
gene            join(51520. .51578,53396. .53583,53762. .53917,56065. .56240)
                /locus_tag="RP1-168L15.4-001"
mRNA            join(51520. .51578,53396. .53583,53762. .53917,56065. .56240)
                /locus_tag="RP1-168L15.4-001"
                /product="novel transcript"
                /note="match: ESTs: AW992696"
gene            join(57352. .57454,65670. .65800,68054. .68179,76558. .77486)
                /gene="RPS6KA2"
                /locus_tag="RP1-168L15.2-004"
mRNA            join(57352. .57454,65670. .65800,68054. .68179,76558. .77486)
                /gene="RPS6KA2"
                /locus_tag="RP1-168L15.2-004"
                /product="ribosomal protein S6 kinase, 90kDa, polypeptide
                2"
                /note="match: ESTs: AL692020 BG195728"
CDS             join(complement(AL159163.40:82039. .82101),
                complement(AL159163.40:78131. .78190),
                AL023775.1:11157. .11207,Z98049.1:113954. .114070,
                Z98049.1:121425. .121506,6546. .6626,8621. .8700,
                12291. .12397,15979. .16016,18253. .18395,26162. .26232,
                27982. .28070,46998. .47062,57352. .57454,65670. .65800,
                68054. .68179,84413. .84502,86292. .86450,93486. .93647,
                96945. .97021,98561. .98678,102972. .103109,104016. .>104140)
                /gene="RPS6KA2"
                /locus_tag="RP1-168L15.2-001"
                /standard_name="OTTHUMP00000017603"
                /codon_start=1
                /product="ribosomal protein S6 kinase, 90kDa, polypeptide
                2"
                /protein_id="CAI19652.1"
                /db_xref="GI:56202586"
                /db_xref="GOA:Q5T8Q1"
                /db_xref="InterPro:IPR000719"
                /db_xref="InterPro:IPR000961"
                /db_xref="InterPro:IPR001245"
                /db_xref="InterPro:IPR002290"
                /db_xref="InterPro:IPR008271"
                /db_xref="UniProt/TrEMBL:Q5T8Q1"
                /translation="MPIAQLLELWKKIEVEPMEIETTEEDLNLDVEPTTEDTAEEGKS
                DSAACKTKVAGSVEEEGVVKEIDISHHVKEGFEKADPSQFELLKVLGQGSYGKVFLVR
                KVKGSDAGQLYAMKVLKKATLKVRDRVRSKMERDILAEVNHPFIVKLHYAFQTEGKLY
                LILDFLRGGDLFTRLSKEVMFTEEDVKFYLAELALALDHLHSLGIIYRDLKPENILLD
                EEGHIKITDFGLSKEAIDHDKRAYSFCGTIEYMAPEVVNRRGHTQSADWWSFGVLMFE
                MLTGSLPFQGKDRKETMALILKAKLGMPQFLSGEAQSLLRALFKRNPCNRLGAGIDGV
                EEIKRHPFFVTIDWNTLYRKEIKPPFKPAVGRPEDTFHFDPEFTARTPTDSPGVPPSA
                NAHHLFRGFSFVASSLIQEPSQQDLHKVPVHPIVQQLHGNNIHFTDGYEIKEDIGVGS
                YSVCKRCVHKATDTEYAVKIIDKSKRDPSEEIEILLRYGQHPNIITLKDVYDDGKFVY
                LVMELMRGGELLDRILRQRYFSEREASDVLCTITKTMDYLHSQGVVHRDLKPSNILYR
                DESGSPESIRVCDFGFAKQLRAGNGLLMTPCYTANFVAPEVLKRQGYDAACDIWSLGI
                LLYTMLAGFTPFANGPDDTPEEILARIGSGKYALSGGNWDSISDAAKDVVSKMLHVDP
                HQRLTAMQVLKHPWVVNREYLSPNQLSRQDVHLVKGAMAATYFALNRTPQAPRLEPVL
                SSNLAQRRGMKRLTSTRL"
```

Query Match 25.7%; Score 514.4; DB 8; Length 178098;
Best Local Similarity 77.4%; Pred. No. 4.4e-71;
Matches 740; Conservative 0; Mismatches 176; Indels 40; Gaps 8;

```
Qy      1071 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 1130
             | || | |||||| |||||||| ||| |||||||| |  ||| |||| ||| |||||| |
Db    167967 ATAAATTTTTTGGTTTCCCAGTACATGTAAAAGTTATAATTACACTATACTATAGTCTAT 167908

Qy      1131 TAAGTGTGCAATAGTGTTATGTCTAAAAAAA------CACATACCTTAATTTTAAAATGC 1184
             |||| |||||||||  ||||||||  |||||        ||| |||||||||  ||||| |
Db    167907 TAAGGGTGCAATAGCATTATGTCTGAAAAAAACAATGGTCATGCCTTAATTTAAAAATAC 167848

Qy      1185 TTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCT 1244
             |||||||||||||||||||| ||||||| ||| | ||||||||||  |||| ||||||||
Db    167847 TTTATTACTAAAAAATGCTAGCAATCATCTGAACCTTCAGTGAGTCATAATATTTTTGCT 167788

Qy      1245 GGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CGGGGGTCAGGTGCTGAA 1293
              ||||| || |||  ||   || |||  | |           ||  |   | |||||||
Db    167787 AGTGGAGGGCCTTGCCTGGATGTTGATGGCTGCTGACTGATCAGGATGGTGGTTGCTGAA 167728

Qy      1294 GCTTAGGGTGGCTGTGGCAGTTTCT-----TAAAACAACAGTGAAGATTGCAATATCAGT 1348
             | ||||||||||||  |||  |||||     ||| ||||||||||| || |   ||   |
Db    167727 GGTTAGGGTGGCTGCAGCAATTTCTTAACATAAGACAACAGTGAAGTTTCCTGCATTGAT 167668

Qy      1349 TGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTA 1408
             |||||||||||||||||||| ||||||| |||  ||| ||||| ||||||||||||||||
Db    167667 TGACTCTTCCTTTCATGAAATATTTCTCTATAGCATGTAATGCTGTTTGATAGCATTTTA 167608

Qy      1409 TGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAA 1467
              |||||||||| || ||| | ||||| | ||| |||||| |||||||        || |
Db    167607 CCCACAGTAGAATTTATTTCAGAATTGAAGTCAGTCCTCTTAAACCCTGTCACTTCTTTA 167548

Qy      1468 CAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTG 1527
             | ||||||| |   |||||||| ||||| ||||||||||||||||||| ||||||
Db    167547 TTAACTAAGTTCACGCAATATTCTAAATCCTTTGTTGTCATTTCAACAATGTTCACAACA 167488

Qy      1528 TCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAG 1587
             |||||||||| |||||||  |||||        | |    |||||| |||| | ||||
Db    167487 TCTTCACCAGGGATAGATTCTGTCTCAAGAAACCACTT----TCTTTGTTCATTCGTAAG 167432

Qy      1588 AAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTC 1647
             ||| ||  |||||||||||||||||| |||||||||| |||||| ||||||   | |||
Db    167431 AAGCAACCCCTCATCTGTTCAAGTTTTCTCATGAGATTGGAGCAATTCAGTCACATTTTC 167372

Qy      1648 AGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCT 1707
             |||      || || |||||||| |||||||||| ||||||||||||||  ||| ||
Db    167371 AGG-----CTGCATTTTTAATTCTAGTTCTCTTGCTATTTCTACCACATCTGCAGTTACT 167317

Qy      1708 TCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTC 1767
             |||| || ||||| |||||||| ||| |||||| |||||||||||||||| ||||||||
Db    167316 TCCTTCACTGAAGGCTTGAACCCCTCAAAGTCACCCATGAGGGTTGGAATCAACTTCTTC 167257

Qy      1768 CAAATTCCTGTTAATATTTATATTTTGACCTCC---CATGAATCATGAATGTTCTTAATG 1824
             |||||||||||||| || || |||||||||||   ||||||| |||||||| ||||||
```

```
Db     167256 CAAATTCCTGTTAATGTTGATGTTTTGACCTCCTCTCATGAATGATGAATGTTTTTAATG 167197

Qy       1825 GCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCC 1884
              || ||  |||||| | ||||||||| |||||||||||||||||| | ||  ||||||

Db     167196 ACATCTAAAATGGTCAGTCCTTTCCAGAAGGTTTTCAATTTACTTTGCCCCAATCCAT-- 167139

Qy       1885 ATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAAT 1944
              |||||||  |||| || ||| ||  | ||| |||||  || |||||||| ||||||

Db     167138 ---CAGAGGAATCACTGTCTATGGCAACTGTAGGCTTATAAAAGGTATTTCTTAAATAAT 167082

Qy       1945 AAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              | | |||||||||  |||| ||||||||||||||   ||||| ||| | |||||||

Db     167081 ATGTCTTGAAAGTCTAAATGACTCCTTGATCCATAAGCTGCAGAATGGGTGTTGTG 167026


RESULT 26
AL357332
LOCUS       AL357332               73991 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-96C13 on chromosome 6, complete
            sequence.
ACCESSION   AL357332
VERSION     AL357332.11  GI:14270145
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 73991)
  AUTHORS   Lovell,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 31, 2001 this sequence version replaced gi:13157565.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-96C13 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .73991
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-96C13"
                     /clone_lib="RPCI-11.1"
     misc_feature    100
                     /note="Clone_right_end: RP1-224E15"
     misc_feature    11941. .11972
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    73892
                     /note="Clone_left_end: RP11-100A16"
ORIGIN

  Query Match             25.6%;  Score 511.2;  DB 8;  Length 73991;
  Best Local Similarity   63.6%;  Pred. No. 1.9e-70;
  Matches 1181;  Conservative    0;  Mismatches  563;  Indels  112;  Gaps   22;

Qy        199 TTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAGAGCTTTATTATTAGGGGC 258
              | |||   |   |  | |  | ||||||  || | | |  ||||| | |||||| |
Db      70258 TCCTTCTTGCATTTATAGCCATGTTTGCTTCATATATTTTGGAGCTGTTTTATTATGATG 70317

Qy        259 ACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCTGTCATTATGAATGTCTGT 318
              | |   | | |   |  |  ||    ||| || |||   | |  |||||||||
Db      70318 ATTTTATTATGATGTCCTCTTG----ATTAATTGATGCTTTTAACATTATGAAATAGCAC 70373

Qy        319 ATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTCTGAATGTGTCTGCTATTAAC 378
              ||   |  ||| || ||| |||| |||| ||| |      | | |||||  ||||||
Db      70374 TTTTTATTACTGGTAATATATCTTGTTCTGAATCT-------ACTCTGTCTAATATTAAT 70426

Qy        379 ATAGCCACTCT-GGCTTTTTAAAATTAGTATTTTTATGGTATATATTTTTCCTTTTTTTT 437
              |||  ||   | || || ||   |||||||||    || ||| | |||| |  || |||
Db      70427 ATAATCATGGTAGGTTTATTTTGATTAGTATTAGCGTGTTATGTCTTTTCCAATTATTTC 70486

Qy        438 TTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGGGTGTCTTATAGGCAGCA 497
              ||||||  || |   ||          || ||||| |||||||  ||  |  | | |  ||
Db      70487 ACTTTTAATTTATGCGTG-------TCTTTATATTTAAAGTTGGATTTGTCAATGTTCCA 70539

Qy        498 TATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAACCTTTTTGTTGGAGTGTT 557
              || |    | ||||  |      ||      || || |    | |||  ||||  ||||
Db      70540 TACACTGTGTTCTTATTAATACGTT-------ATGATATATCACCTTTAATTGGGATGTT 70592

Qy        558 TAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTTGCTATACCATCTTTTCA 617
              || |||||||||||||  |||| || |  | ||| |  | |    | |||||||||
Db      70593 CAGTCCATTTACATTTAATGTACTTTTTAATATGATGGGTTACAGTCTACCATCTTCCTC 70652

Qy        618 TTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCATCTTTGACCATTTTCTTT 677
              || ||||| | | |     || ||      |   ||||  |  | |||  || | ||
Db      70653 TTAGTTTTCTGTTTCCCAAATATTCATTATTCCCTTTAATTCTTTTTTTCCTTATTAGGA 70712

Qy        678 AGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTTTTTAGTTATACCTCTTA 737
              | || |  | | ||||| |  |   |||| |  ||     ||| ||   || | |
Db      70713 ATTAATTGAGAATTTTTGTAATTCCTTTTATCTTCTTTACTAACTTATTT---GCTATAA 70769

Qy        738 AAATTTTTTTTTCTGTTTATGTAGGATTTATAATATACATCTTTAACTTATCACAGATT 797
              |||  |  ||  |   || | ||  ||  | |||||||||||| || ||||| |||  |
Db      70770 CTCTTTATCTTAGTACCTTCTTTAATATACACAATATACATCTTAAAATTATCTCAGTCT 70829

Qy        798 ACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAACCTTACAAGAGTATATTT 857
              |||||||||  |||  ||||| ||||  ||||      |  || ||||||| ||||||  |
Db      70830 TCCTTCAAATCATATGATACCACCTCACTTGTAGCACAAGAAACTTACAACAGTATA-CT 70888

Qy        858 TCATTTCTGTCTCCTAATTTTTATGCTATGTCTATAATACATTAGGTTTGTTGTTGTTTG 917
              ||||||    ||||    ||| ||||||   ||   | ||||    || |  || | |
Db      70889 CCATTTCACCTTCCTGGCCTTTGTGCTATTGTTACTGTGCATTTTACTTCTATTTATGTA 70948

Qy        918 TTT--------------------TTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATT 956
              |                    | |  ||| |   ||     |    ||   | | |||
Db      70949 ATAAATCACACAGTATATTATTACTAATTTTACTTTAGTCAATTATCTTCTAAATATATT 71008

Qy        957 TTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATA 1016
              |  |||        |  ||||| ||||| |  |||||| | ||||||  |||  | |
Db      71009 TAAATAAAAACATATTTTTATACAGCCATACTTGAGAGATATTCTGGGTTCTGTTTAAGA 71068

Qy       1017 CCACCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGT 1076
              ||||  |||||               ||||  ||||| |||||||||||||||||| | ||
Db      71069 CCACTGCAATA--------------AAGTAAATATCTCAATAAAGTGAGTCACATATGT 71113

Qy       1077 CTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTG 1136
              || | | || |||| |||||||||||| || ||| |||| |||| ||||| |||||||
Db      71114 TTTCTTG-TTTCCAGGGCATATAAAAGTTATGTTTACACTATACTG-AGTCTATTAAGTG 71171
```

```
Qy       1137 TGCAATAGTGTTATGTCTAAAAAAACA-----CATACCTTAATTTTAAAATGCTTTATTA 1191
             ||||||| |||||  || || |||||     || | |||| ||| |||||  ||||||
Db      71172 TGCAATATCATTATGACTTAATAAACAATATGCACATCTTATTTTGAAAATACTTTATGG 71231

Qy       1192 CTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAA 1251
             |||||||  |||||  ||  | |||||  |||||  |||  ||||||||||||| |||||
Db      71232 CTAAAAATCACTAACCATAGTCTGAGCTTTCAGCTAGTCATAATCTTTTTGCTAGTGGAG 71291

Qy       1252 GGTCTTTTCTTATTGATGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGG 1300
             ||||||  |||  || |||  | |             ||  |     |||||||  ||
Db      71292 GGTCTTGCCTTGATGTTGATGGCTGCTAACTCATCAGGATGGTGATTGCTGAAAGTTTAT 71351

Qy       1301 GTGGCTGTGGCAGTTTCTTAAAACAACA----GTGAAGATTGCAATATCAGTTGACTCTT 1356
             |  | |||| || ||||||||||| || |     ||| | || | |       |||||||||
Db      71352 GCAGATGTGCCAATTTCTTAAAATAAGAAAACATGATGTTTCCCACGGTGATTGACTCTT 71411

Qy       1357 CCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGT 1416
              ||||||||||  ||||||| |||  ||  |  |||||||| | |||||   ||||||
Db      71412 TCTTTCATGAAAAGTTTCTCTGTAGCATGCAACACTTTTTGATGGTATTTTCCCCACAGT 71471

Qy       1417 AGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAA 1475
             |||||||||| ||| || || |||||||||| |||| ||||| |  ||| |  | ||||
Db      71472 AGAACTTCTTTCAAATTTAGAGTCAATCCTCTAAAACTCTGCTGTTGTTTTATCAGCTAA 71531

Qy       1476 GTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACC 1535
             ||| || |||||||| ||| | | ||||||||||||||||| |||||||  |||||| |
Db      71532 GTTTATGTAATATTCTAAATACTTGGTTGTCATTTCAACAATGTTCACAGCATCTTCATC 71591

Qy       1536 AGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATT 1595
             ||||| |||||||||| || | |   | |    |||||||||| ||||||| | || |
Db      71592 AGGAGCAGATTCCATCACAATACACCACTT----TCTTTGCTCATTCATAAGAGGCAACT 71647

Qy       1596 CCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCA 1655
             ||||| ||||| || ||||||||||||||  |||||| ||| ||  ||||||||
Db      71648 CCTCACCTGTTAAAATTTTATCATGAGATTAAAGCAATTCAGACACATCTTCAGGC---- 71703

Qy       1656 CTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCAT 1715
                ||| |  |||| |||||||| || |||| ||||||| ||| |||  |||| |||
Db      71704 ---CACCTCAAATTATAGTTCTCTTACTATTTCCACCACATTTGTAGTTAATTCCCTCAT 71760

Qy       1716 TGAAGTCTTGAA-CCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTC 1774
              |||||||||| ||| || || | ||  |||||||||| |||| ||| |||| ||| ||
Db      71761 GGAAGTCTTGAAGCCTGTCAAAATAATTTATGAGGGTTGAAATCAACTCCTTCTAAACTC 71820

Qy       1775 CTGTTAATATTTATATTTTGACCTC-CCATGAATCATGAATGTTCTTAATGGCACCTGGA 1833
             || ||||| || ||||| | |||||  |   ||||||||||||   |||||||  | ||
Db      71821 CTATTAATGTTCATATTCTCACCTCTTCCCAAATCATGAATGTTGCCAATGGCATTTAGA 71880

Qy       1834 ATGGTGAAT-CCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAG 1892
              ||| | ||| ||||| || | ||| ||| |||||||| |  |||||||||  |   | .
Db      71881 ATGATTAATCCCCTTTTCAGAGGGTGTTCCATTTACTTTGCTCAGATCCATCAAAGAAACC 71940

Qy       1893 GAT--------CCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAAT 1944
                |          || |  | |||     |   | ||||  |||||||||||| ||||||
Db      71941 ACTATCTATGGGAGCTATGTAAGCCTTACAATACATTATAAAATGTATTTCTTAAATAAT 72000

Qy       1945 AAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              ||| ||| |||||  |||| ||||||||||||||   ||||| | | |||||||||
Db      72001 AAGACTT-AAAGTCAAAATGACTCCTTGATCCATGGGCTGCAGATTGGATGTTGTG 72055
```

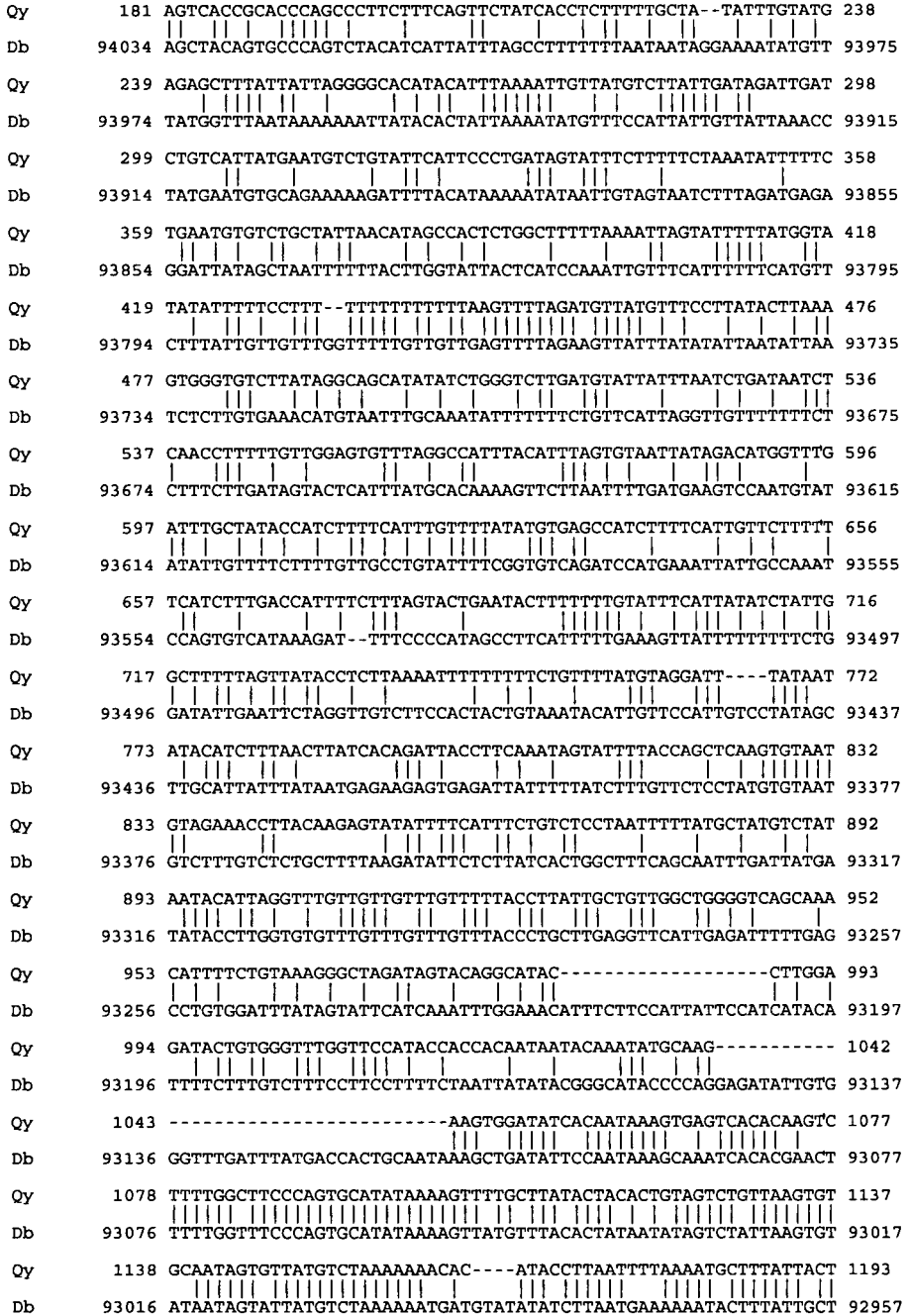
```
RESULT 27
AL590674/c
LOCUS       AL590674              114056 bp    DNA     linear   PRI 24-OCT-2001
DEFINITION  Human DNA sequence from clone RP11-286N24 on chromosome 1, complete
            sequence.
ACCESSION   AL590674
VERSION     AL590674.5  GI:16444743
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 114056)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 25, 2001 this sequence version replaced gi:13990635.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-286N24 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-286N24 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-286N24 is at 114056 in this
            sequence. The true left end of clone RP11-2L13 is at 724 in this
            sequence. The true right end of clone RP11-152M20 is at 2000 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .114056
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-286N24"
                     /clone_lib="RPCI-11.1"
     misc_feature    6238. .6251
                     /note="Sequence from overlapping clone bA2L313 (AL596266).
                     Assembly confirmed by restriction digest."
     misc_feature    7035. .7250
                     /note="Sequence from overlapping clone bA2L313 (AL596266).
                     Assembly confirmed by restriction digest."
ORIGIN

  Query Match             25.6%;  Score 511.2;  DB 8;  Length 114056;
  Best Local Similarity   58.3%;  Pred. No. 1.6e-70;
  Matches 1218;  Conservative    0;  Mismatches  758;  Indels  112;  Gaps   14;

Qy          3 ATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCAGC 62
              |||||||||| ||||| | |  |||||||| | |||||||  || |||| || | |||||
Db      94214 ATTCTCCTGCTTCAGCTTTCTGAGTAGCTGGGATTACAGGCACCCGCCATCATGCCCAGC 94155

Qy         63 TAATTTTTGTA--TTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120
              ||||||||||  |||||||||||| ||| ||||||| |||| ||||||||||| |||||
Db      94154 TAATTTTTGTATTTTTTAGTAGAGATGGGATTTCACCATGTTGGCCAGGCTGGTCTCAAA 94095

Qy        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180
              |||||||||||| || |||  ||   |||||||||||||| |||| |||||| | |  ||
Db      94094 CTCCTGACCTCAGGTAATCCACCCATCTCAGCCTCCCAAAGTGCTGGGATTATAAGCATG 94035
```



RESULT 28
AL596266/c
LOCUS AL596266 128479 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-2L13 on chromosome 1 Contains a translational inhibitor protein p14.5 (UK114) pseudogene, complete sequence.
ACCESSION AL596266
VERSION AL596266.3 GI:15626484
KEYWORDS HTG; UK114.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 128479)
AUTHORS Chapman,J.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 16, 2001 this sequence version replaced gi:15028804.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-2L13 is from the library RPCI-11.1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .128479
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-2L13"
                     /clone_lib="RPCI-11.1"
     misc_feature    1
                     /note="Clone_left_end: RP11-2L13"
     gene            75806. .76205
                     /locus_tag="RP11-2L13.1-001"
                     /pseudo
     CDS             75806. .76205
                     /locus_tag="RP11-2L13.1-001"
                     /note="match: proteins: Sw:P52758 Sw:P52759 Sw:P52760
                     Sw:P80601 Tr:O97509 Tr:Q9WUV8"
                     /pseudo
                     /codon_start=1
                     /product="translational inhibitor protein p14.5 (UK114)
                     pseudogene"
     misc_feature    113336
                     /note="Clone_right_end: RP11-286N24"
     misc_feature    126480
                     /note="Clone_left_end: RP11-567B20"
ORIGIN

  Query Match             25.6%;  Score 511.2;  DB 8;  Length 128479;
  Best Local Similarity   58.3%;  Pred. No. 1.6e-70;
  Matches 1218;  Conservative    0;  Mismatches  758;  Indels  112;  Gaps   14;

Qy          3 ATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCAGC 62
              ||||||||||  |||||  | |  |||||||| | |||||||   || |||| || | |||||
Db      93497 ATTCTCCTGCTTCAGCTTTCTGAGTAGCTGGGATTACAGGCACCCGCCATCATGCCCAGC 93438

Qy         63 TAATTTTTGTA--TTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120
              ||||||||||   ||||||||||| ||| ||||||| |||| |||||||||||  |||||
Db      93437 TAATTTTTGTATTTTTTAGTAGAGATGGGATTTCACCATGTTGGCCAGGCTGGTCTCAAA 93378

Qy        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180
              ||||||||||| || |||  ||   |||||||||||||| |||| |||||| | |  ||
Db      93377 CTCCTGACCTCAGGTAATCCACCCATCTCAGCCTCCCAAAGTGCTGGGATTATAAGCATG 93318

Qy        181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTA--TATTTGTATG 238
              ||  || |  ||||| |    |      ||     |  || |  ||   | | | |
Db      93317 AGCTACAGTGCCCAGTCTACATCATTATTTAGCCTTTTTTTAATAATAGGAAAATATGTT 93258

Qy        239 AGAGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGAT 298
              | |||| ||  |     | | ||  |||||||   | |   |||||| ||
Db      93257 TATGGTTTAATAAAAAAATTATACACTATTAAAATATGTTTCCATTATTGTTATTAAACC 93198

Qy        299 CTGTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTC 358
              ||    |      |  || |        |||  |||    |           |
Db      93197 TATGAATGTGCAGAAAAAGATTTTACATAAAAATATAATTGTAGTAATCTTTAGATGAGA 93138

Qy        359 TGAATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTA 418
              || | |  ||  | ||   | |  | |    |    ||    |||||   ||
Db      93137 GGATTATAGCTAATTTTTTACTTGGTATTACTCATCCAAATTGTTTCATTTTTTCGTGTT 93078

Qy        419 TATATTTTTCCTTT--TTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAA 476
              |  || |  |||  ||||| || || ||||||||| |||||  | |   |  |  | ||
Db      93077 CTTTATTGTTGTTTGGTTTTTGTTGTTGAGTTTTAGAAGTTATTTATATATTAATATTAA 93018

Qy        477 GTGGGTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCT 536
              |||   | | | |   | |  |   | ||  |||     |    | | |  |||
Db      93017 TCTCTTGTGAAACATGTAATTTGCAAATATTTTTTTCTGTTCATTAGGTTGTTTTTTTCT 92958

Qy        537 CAACCTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTG 596
              |  |||  | |   ||  |  ||      ||| | |   |  || |    |
Db      92957 CTTTCTTGATAGTACTCATTTATGCACAAAAGTTCTTAATTTTGATGAAGTCCAATGTAT 92898

Qy        597 ATTTGCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTT 656
              ||  |  | | |  |  || | | | ||||   ||| ||      |    | | |    |
Db      92897 ATATTGTTTTCTTTTGTTGCCTGTATTTTCGGTGTCAGATCCATGAAATTATTGCCAAAT 92838

Qy        657 TCATCTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTG 716
              ||   |     |  | |||     |         ||||||  |  | ||| | | |  ||
Db      92837 CCAGTGTCATAAAGAT--TTTCCCCATAGCCTTCATTTTTGAAAGTTATTTTTTTTTCTG 92780

Qy        717 GCTTTTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATT----TATAAT 772
              | | || | || ||  |  |       |  | |   |    |||   |||    ||||
Db      92779 GATATTGAATTCTAGGTTGTCTTCCACTACTGTAAATACATTGTTCCATTGTCCTATAGC 92720

Qy        773 ATACATCTTTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAAT 832
              | |||  || |         ||| |    | |  |     |||      |  | |||||||
Db      92719 TTGCATTATTTATAATGAGAAGAGTGAGATTATTTTTATCTTTGTTCTCCTATGTGTAAT 92660

Qy        833 GTAGAAACCTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATGTCTAT 892
              ||      ||          | || |||  || |  ||      |   |   | | |
Db      92659 GTCTTTGTCTCTGCTTTTAAGATATTCTCTTATCACTGGCTTTCAGCAATTTGATTATGA 92600

Qy        893 AATACATTAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAA 952
              |||| || |  |  |||||  ||  |||  |||  |||  |||  || | |     |
Db      92599 TATACCTTGGTGTGTTTGTTTGTTTGTTTACCCTGCTTGAGGTTCATTGAGATTTTTGAG 92540

Qy        953 CATTTTCTGTAAAGGGCTAGATAGTACAGGCATAC------------------CTTGGA 993
              | | |    | |  | |  ||   |   | | ||                    | | |
Db      92539 CCTGTGGATTTATAGTATTCATCAAATTTGGAAACATTTCTTCCATTATTCCATCATACA 92480

Qy        994 GATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAAG---------- 1042
              | || ||  |||  |||| |  |      |   |     |||  | ||
Db      92479 TTTTCTTTGTCTTTCCTTCCTTTTCTAATTATATACGGGCATACCCCAGGAGATATTGTG 92420

Qy       1043 -------------------------AAGTGGATATCACAATAAAGTGAGTCACACAAGTC 1077
                                       |||  |||||  ||||||||  | ||||||  |
Db      92419 GGTTTGATTTATGACCACTGCAATAAAGCTGATATTCCAATAAAGCAAATCACACGAACT 92360

Qy       1078 TTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGT 1137
              |||||| |||||||||||||||||||||| || ||| |||| | | |||||| ||||||||
Db      92359 TTTTGGTTTCCCAGTGCATATAAAAGTTATGTTTACACTATAATATAGTCTATTAAGTGT 92300

Qy       1138 GCAATAGTGTTATGTCTAAAAAAACAC----ATACCTTAATTTTAAAATGCTTTATTACT 1193
              ||||||  |||||||||||||||  |    ||| ||||||   |||||  |||||||  ||
Db      92299 ATAATAGTATTATGTCTAAAAAATGATGTATATATCTTAATGAAAAAATACTTTATTGCT 92240
```

```
Qy       1194 AAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGG 1253
             |||| ||||||| ||| ||   ||| |||||  |||  ||||||||||||||||||  |
Db      92239 AAAATATGCTAAGAATTATCCAAGCCTTCAGAAAGTCATAATCTTTTTGCTGGTGGGGAG 92180

Qy       1254 TCTTTTCTTATTGATGACTG----------ATCGGGGGTCAGGTGCTGAAGCTTAGGGT 1302
             ||||   | | || |||  |             |||     | |||||||| || ||||
Db      92179 TCTTGCTTCAATGTTGATGGCAACTGACTGATCAGGGTAGTGGCTGCTGAAGGTTTGGGT 92120

Qy       1303 GGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATTGCAATATCA----GTTGACTCTTCC 1358
             |||||||||||||| | |||| || |   ||        |  ||    |||||||||| |
Db      92119 GGCTGTGGCAGTTTTTAAAAATAAGACAAAAATGAAGTTTTCCACATTGTTGACTCTTTC 92060

Qy       1359 TTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAG 1418
             |||||    |||||||||| ||| ||  |||||  | |||||| ||||||  || |||||
Db      92059 TTTCACCGTAGATTTCTCTGTAGCATGCAATGCTGCTCGATAGTATTTTACTCATAGTAG 92000

Qy       1419 AACTTCTTTGAAAATTGGATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTT 1478
             |||||||||  |  ||||||   |  ||||||||||||||     || |  | ||||||
Db      91999 AACTTCTTTCATTATTGGAGTCA-ATTCTCAAACCCTGCTGCTGCTTTATCAACTAAGTG 91941

Qy       1479 AATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGG 1538
               ||  ||| |||| |||| |||||||  |||||||| | |||||||  | ||| ||||
Db      91940 TATGGAATCTTCTAAATCCTTTGTTGTTATTTCAACAGTGTTCACAGCCTTTTCGCCAGA 91881

Qy       1539 AGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCAT------CCATAAGAAGAA 1592
             ||||||||  | || ||         | |   |   ||     |||       ||||||||||
Db      91880 AGTAGATTGCTTCACAAGAAACCCTAAGACACTTTGCTCCATAAGATACCATAAGAAGCC 91821

Qy       1593 ATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCC 1652
             ||||||||||  ||||| |||||||||| ||||||||||||| |||||   |||||||||
Db      91820 ATTCCTCATCCATTCAAATTTTATCATTAGATTGCAGCAATTCAGTCTCATCTTCAGGC- 91762

Qy       1653 TCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTC 1712
                  | || ||| |||||||||||||||||| |||||||  ||   |    ||| |||
Db      91761 ----------TCTACTTCTAGTTCTCTTGCTGTTTCCACCACATTTGCAGAGGCTTTCTC 91712

Qy       1713 CATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAAT 1772
             || ||| |||||||||| ||| |||||||||| |||| |||||||| ||||||| ||||
Db      91711 CACTGAGGTCTTGAACCCCTCAAAGTCATCCACGAGGATTGGAATCAACTTCTTTCAAAC 91652

Qy       1773 TCCTGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGG 1832
             |||  ||||| || || |||||| | || ||||||||   |||||||||||| || || |
Db      91651 TCCCATTAATGTTGATCTTTTGATCGCCTATGAATCACAGATGTTCTTAATGACATCTAG 91592

Qy       1833 AATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAG 1892
             |||| ||||| ||||||| |||||||||||||| |||| | | |||||||||
Db      91591 AATGATGAATTCTTTCCAGAAGGTTTTCAATTAACTTTGCCTAGATCCATC--------A 91540

Qy       1893 GATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTG 1952
             ||   |      |  || | | |||| ||||  | |||||||| | |||||||||  |||
Db      91539 GAGAAATCACTATGGCAACTAATAGTCTTACAGCATGTATTTCCTAAATAATAAGATTTG 91480

Qy       1953 AAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
             | |||||||||||||||||||| |||     | || ||| ||| |||||
Db      91479 ATAGTTGAAATTACTCCTTGATTCATGGGTTACAGAATGGATTTTGTG 91432
```

RESULT 29
HSDJ45N11/c

```
LOCUS       HSDJ45N11              99025 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP1-45N11 on chromosome 6q15-16.1,
            complete sequence.
ACCESSION   AL132996
VERSION     AL132996.4  GI:8648910
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 99025)
  AUTHORS   Collier,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 21, 2000 this sequence version replaced gi:7940170.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            RP1-45N11 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .99025
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704N1145"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q15-16.1"
                     /clone="RP1-45N11"
                     /clone_lib="RPCI-1"
     misc_feature    1
                     /note="Clone_left_end: RP1-45N11"
     misc_feature    99025
                     /note="Clone_right_end: RP1-45N11"
ORIGIN
```

```
  Query Match             25.6%;  Score 511;  DB 8;  Length 99025;
  Best Local Similarity   75.2%;  Pred. No. 1.8e-70;
  Matches  786;  Conservative    0;  Mismatches  210;  Indels   49;  Gaps   10;

Qy        975 AGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAA 1034
              |||||  | ||||||  |||||| ||  ||||||||||||    || | || ||
Db      75033 AGTACGTGTATACCTCTGAGATATTGCAGGTTTGGTTCCAGGTCATCTCAGTA------- 74981

Qy       1035 TATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGC 1094
                || || | ||||||||||||  ||||||  || | | || | || ||||| |
Db      74980 --------AAATGAACATCACAATAAAGCCAGTCACGTAAATTTGTTTGTTTTCCAGTTC 74929

Qy       1095 ATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATA--GTGTTATGT 1152
              |||||||||| ||  ||||||| ||| |||| | |||||||||| |||    | ||| |
Db      74928 ATATAAAAGTTATGTGTATACTAAACTATAGTTTATTAAGTGTGCCATAACATTGTATCT 74869

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212
              ||||||   |||||||||| |  |||||  |||||| |||||||||| |||||||||
Db      74868 TTAAAAATGTACATACCTTAAATAAAAAATAGTTTATTGCTAAAAAATGCCAACAATCAT 74809

Qy       1213 TTGAGCATTCAGTGAGTTGTAA--TCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGA 1270
              ||||| |||||||||| ||||  | |||||||||||||| || ||    ||   || |||
Db      74808 CTGAGCCTTCAGTGAGTCGTAATTTTTTTTTGCTGGTGGAGGGCCTCGCCTCGATGTTGA 74749

Qy       1271 CTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT 1319
                | |            ||| |   | |||||||| || || ||||||||||  |||||
Db      74748 TGGCTGCTGACTCATCAGGGTGGTGGTTGCTGAAGGTTGGGATGGCTGTGGCAACTTCTT 74689
```

```
Qy       1320 AAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
             |||| | |   |  |||      |||| | ||||||| |||||||||||||||||| || |
Db      74688 AAAATCAGA-AAATAATTAATGGATCAATGGACTCTTTCTTTCATGAAAGATTTCACTGT 74630

Qy       1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-T 1438
             ||  ||  ||||| ||||| |||||| ||  || |||| ||||||||| ||||||| | |
Db      74629 AGCATGCAATGCTGTTTGACAGCATTATACCCATAGTAAAACTTCTTTCAAAATTGTAGT 74570

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
             |||||||||| |||||||||     || |  |  |||||| ||||| |||||| | |||
Db      74569 CAATCCTCTCCAACCCTGCTGCTGCTTTATCAAGTAAGTTTATATAGTATTCTAACTCCT 74510

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
             ||||||||||||||||||  |||||||   ||||| ||||||||||||||||||||||
Db      74509 TTGTTGTCATTTCAACAACATTCACAGCACCTTCATCAGGAGTAGATTCCATCTCAAGAA 74450

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
                |  |    ||||| |||||||||| ||| || ||||||||||||||  ||||||
Db      74449 ACCACCT----TCTTTACTCATCCATAAAAAGCAACTCCTCATCTGTTCAAACTTTATCT 74394

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
             |||||||||||||||  |  | | |||||||||     ||||||||  ||||| | || ||
Db      74393 TGAGATTGCAGCAATTCTCTGACGTCTTCAGG-----CTTCACTTCGAATTCTACTTATC 74339

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
             ||||| |||| | |||| |||  ||| |||| |||| |  ||| ||||||| ||| ||||
Db      74338 TTGCTATTTCCATCACAACTGCAGTTACTTCTTCCACTTCAGTATTGAACCCCTCAAAGT 74279

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC-- 1796
             ||| ||||||| |||||||| |||||||||| | ||||| |||  || ||||||||||
Db      74278 CATTCATGAGGTTTGGAATCAACTTCTTCCATACTCCTGCTAACGTTGATATTTTGACCT 74219

Qy       1797 -CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855
              |||||||||||||  |||||||||||||||| || ||||| ||||| ||||||  ||||
Db      74218 GCTCCCATGAATCACAAATGTTCTTAATGGCATCTAGAATGATGAATTCTTTCCGGAAGG 74159

Qy       1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915
             |||||||||||||| | |  |||||||      ||||||  |||| || ||| ||| |||
Db      74158 TTTTCAATTTACTTTGCCACGATCCAT-----TAGAGGAATCACTATCTATGGCAGCTAT 74104

Qy       1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
             ||| ||||| |||||||||||| ||||| ||| ||||||||||||||||||| ||| |
Db      74103 AGCTTTATGAAATGTATTTCTTAAATAAGAAGTCTTGAAAGTTGAAATTACTCTTTGGTT 74044

Qy       1976 CATTTTCTGCAAAATAGATGTTGTG 2000
              |||   ||||| ||||||||||||
Db      74043 CATGGGCTGCAGAATAGATGTTGTG 74019
```

RESULT 30
AC027156
LOCUS AC027156 172061 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 6 clone RP11-450D8 map 6, WORKING DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION AC027156
VERSION AC027156.2 GI:7596815
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172061)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 6, clone RP11-450D8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172061)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 172061)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 19, 2000 this sequence version replaced gi:7331458.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L5972
Center clone name: 450_D_8
-------------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161486 bases at least Q40
Consensus quality: 167277 bases at least Q30
Consensus quality: 169194 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 170261; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1174: contig of 1174 bp in length
          *     1175     1274: gap of 100 bp
          *     1275     3364: contig of 2090 bp in length
          *     3365     3464: gap of 100 bp
          *     3465     5872: contig of 2408 bp in length
          *     5873     5972: gap of 100 bp
          *     5973    10787: contig of 4815 bp in length
          *    10788    10887: gap of 100 bp
          *    10888    13492: contig of 2605 bp in length
          *    13493    13592: gap of 100 bp
          *    13593    17295: contig of 3703 bp in length
          *    17296    17395: gap of 100 bp
          *    17396    22203: contig of 4808 bp in length
          *    22204    22303: gap of 100 bp
          *    22304    27575: contig of 5272 bp in length
          *    27576    27675: gap of 100 bp
          *    27676    31757: contig of 4082 bp in length
          *    31758    31857: gap of 100 bp
          *    31858    37767: contig of 5910 bp in length
          *    37768    37867: gap of 100 bp
          *    37868    42646: contig of 4779 bp in length
          *    42647    42746: gap of 100 bp
          *    42747    50274: contig of 7528 bp in length
          *    50275    50374: gap of 100 bp
          *    50375    59804: contig of 9430 bp in length
          *    59805    59904: gap of 100 bp
          *    59905    69512: contig of 9608 bp in length
          *    69513    69612: gap of 100 bp
          *    69613    80756: contig of 11144 bp in length
          *    80757    80856: gap of 100 bp
          *    80857   100937: contig of 20081 bp in length
          *   100938   101037: gap of 100 bp
          *   101038   118877: contig of 17840 bp in length
          *   118878   118977: gap of 100 bp
          *   118978   144244: contig of 25267 bp in length
          *   144245   144344: gap of 100 bp
          *   144345   172061: contig of 27717 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172061
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6"
                     /clone="RP11-450D8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1174
                     /note="assembly_fragment"
     gap             1175. .1274
                     /estimated_length=100
     misc_feature    1275. .3364
                     /note="assembly_fragment"
     gap             3365. .3464
                     /estimated_length=100
     misc_feature    3465. .5872
                     /note="assembly_fragment"
     gap             5873. .5972
                     /estimated_length=100
     misc_feature    5973. .10787
                     /note="assembly_fragment"
     gap             10788. .10887
                     /estimated_length=100
     misc_feature    10888. .13492
                     /note="assembly_fragment"
     gap             13493. .13592
                     /estimated_length=100
     misc_feature    13593. .17295
                     /note="assembly_fragment"
     gap             17296. .17395
                     /estimated_length=100
     misc_feature    17396. .22203
                     /note="assembly_fragment"
     gap             22204. .22303
                     /estimated_length=100
     misc_feature    22304. .27575
                     /note="assembly_fragment"
     gap             27576. .27675
                     /estimated_length=100
     misc_feature    27676. .31757
                     /note="assembly_fragment"
     gap             31758. .31857
                     /estimated_length=100
     misc_feature    31858. .37767
                     /note="assembly_fragment"
     gap             37768. .37867
                     /estimated_length=100
     misc_feature    37868. .42646
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     gap             42647. .42746
                     /estimated_length=100
     misc_feature    42747. .50274
                     /note="assembly_fragment"
     gap             50275. .50374
                     /estimated_length=100
     misc_feature    50375. .59804
                     /note="assembly_fragment"
     gap             59805. .59904
                     /estimated_length=100
     misc_feature    59905. .69512
                     /note="assembly_fragment"
     gap             69513. .69612
```

Query Match 25.6%; Score 511; DB 14; Length 172061;
Best Local Similarity 75.2%; Pred. No. 1.5e-70;
Matches 786; Conservative 0; Mismatches 210; Indels 49; Gaps 10;

```
Qy       975 AGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAA 1034
             |||||| | ||||||  |||||| ||  ||||||||||||    || | || ||| |
Db     85664 AGTACATGTATACCTCTGAGATATTGCAGGTTTGGTTCCAGGTCATCTCAGTAAAA---- 85719

Qy      1035 TATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGC 1094
                 || | ||||||||||||  ||||||  || | | || | || ||||| |
Db     85720 -----------TGAACATCACAATAAAGCCAGTCACGTAAATTTGTTTGTTTTCCAGTTC 85768

Qy      1095 ATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATA--GTGTTATGT 1152
             |||||||||| ||  ||||||| ||| |||| | |||||||||| |||    |  ||| |
Db     85769 ATATAAAAGTTATGTGTATACTAAACTATAGTTTATTAAGTGTGCCATAACATTGTATCT 85828

Qy      1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212
             ||||||   |||||||||| |  |||||  |||||| |||||||||| |||||||||
Db     85829 TTAAAAATGTACATACCTTAAATAAAAAATAGTTTATTGCTAAAAAATGCCAACAATCAT 85888

Qy      1213 TTGAGCATTCAGTGAGTTGTAA--TCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGA 1270
             ||||| |||||||||| |||| | |||||||||||||| || ||    ||  || |||
Db     85889 CTGAGCCTTCAGTGAGTCGTAATTTTTTTTTGCTGGTGGAGGGCCTCGCCTCGATGTTGA 85948

Qy      1271 CTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT 1319
               | |            ||| |   | |||||||| || || ||||||||||  |||||
Db     85949 TGGCTGCTGACTCATCAGGGTGGTGGTTGCTGAAGGTTGGGATGGCTGTGGCAACTTCTT 86008

Qy      1320 AAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
             ||||  | |   |  |||     |||| | ||||||| ||||||||||||||||| || |
Db     86009 AAAATCAGA-AAATAATTAATGGATCAATGGACTCTTTCTTTCATGAAAGATTTCACTGT 86067

Qy      1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-T 1438
             ||  ||  ||||| ||||| |||||| ||  || |||| |||||||| ||||||| | |
Db     86068 AGCATGCAATGCTGTTTGACAGCATTATACCCATAGTAAAACTTCTTTCAAAATTGTAGT 86127

Qy      1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
             |||||||||| |||||||||      || |  | |||||| ||||| |||||| | |||
Db     86128 CAATCCTCTCCAACCCTGCTGCTGCTTTATCAAGTAAGTTTATATAGTATTCTAACTCCT 86187
```

```
Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
             ||||||||||||||||||  |||||||    ||||| ||||||||||||||||||||
Db      86188 TTGTTGTCATTTCAACAACATTCACAGCACCTTCATCAGGAGTAGATTCCATCTCAAGAA 86247

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
                |  |    ||||| |||||||||||| ||| || ||||||||||||||  ||||||
Db      86248 ACCACCT----TCTTTACTCATCCATAAAAAGCAACTCCTCATCTGTTCAAACTTTATCT 86303

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
             ||||||||||||||  |  |  ||||||||     ||||||||  ||||| | || ||
Db      86304 TGAGATTGCAGCAATTCTCTGACATCTTCAGG-----CTTCACTTCGAATTCTACTTATC 86358

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
             |||||  |||| | |||| |||  ||| |||| |||| |  ||| ||||||| ||| ||||
Db      86359 TTGCTATTTCCATCACAACTGCAGTTACTTCTTCCACTTCAGTATTGAACCCCTCAAAGT 86418

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC-- 1796
             |||  |||||||  ||||||||  ||||||||||  | |||||  |||   ||  ||||||||||
Db      86419 CATTCATGAGGTTTGGAATCAACTTCTTCCATACTCCTGCTAACGTTGATATTTTGACCT 86478

Qy       1797 -CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855
              ||||||||||||  |||||||||||||||| || |||||  |||||  ||||||  ||||
Db      86479 GCTCCCATGAATCACAAATGTTCTTAATGGCATCTAGAATGATGAATTCTTTCCGGAAGG 86538

Qy       1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915
             |||||||||||||  | |  |||||||       ||||||  ||||  ||  |||  ||| |||
Db      86539 TTTTCAATTTACTTTGCCACGATCCAT-----TAGAGGAATCACTATCTATGGCAGCTAT 86593

Qy       1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
             |||  |||||  ||||||||||||  |||||  |||  ||||||||||||||||||||  ||| |
Db      86594 AGCTTTATGAAATGTATTTCTTAAATAAGAAGTCTTGAAAGTTGAAATTACTCTTTGGTT 86653

Qy       1976 CATTTTCTGCAAAATAGATGTTGTG 2000
              |||   |||||  ||||||||||||||
Db      86654 CATGGGCTGCAGAATAGATGTTGTG 86678
```

```
RESULT 31
HS370M22
LOCUS       HS370M22              143747 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP3-370M22 on chromosome 22, complete
            sequence.
ACCESSION   Z82206
VERSION     Z82206.1  GI:2677628
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 143747)
  AUTHORS   Burgess,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 12, 1997 this sequence version replaced gi:1772957.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            RP3-370M22 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .143747
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704M22370"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="RP3-370M22"
                     /clone_lib="RPCI-3"
     gene            complement(join(3536. .4154,7268. .7510,32902. .33405,
                     39197. .39292))
                     /locus_tag="RP3-370M22.3-003"
     mRNA            complement(join(3536. .4154,7268. .7510,32902. .33405,
                     39197. .39292))
                     /locus_tag="RP3-370M22.3-003"
                     /note="match: cDNAs: AF062084"
     CDS             complement(join(4093. .4154,7268. .7510,32902. .33250))
                     /locus_tag="RP3-370M22.3-003"
                     /standard_name="OTTHUMP00000028839"
                     /codon_start=1
                     /protein_id="CAI42712.1"
                     /db_xref="GI:57209987"
                     /db_xref="UniProt/TrEMBL:Q5H9F7"
                     /translation="MAFRRQVKNFVKNYSDAEIKVREATSNDPWGPSSSLMLDISDLT
                     FNTISLSEIMNMLWHRLNDHGKNWRHVYKSLTLMDYLIKNGSKKVIQHCREGFCNLQT
                     LKDFQHIDEAGKDQGYYIREKSKQVITLLMDEPLLCKEREVACRTRQRTSHSILFSKR
                     QLGSSNSLTACTSAPTPDISASEKKYKLPKFGRLHNKTHGSRSDFNFQVLLLKKYIS"
     gene            <20799. .21898
                     /locus_tag="RP3-370M22.2-001"
     mRNA            <20799. .21898
                     /locus_tag="RP3-370M22.2-001"
     CDS             <20799. .21617
                     /locus_tag="RP3-370M22.2-001"
                     /standard_name="OTTHUMP00000028838"
                     /codon_start=1
                     /protein_id="CAI42711.1"
                     /db_xref="GI:57209986"
                     /db_xref="GOA:Q5H9F8"
                     /db_xref="InterPro:IPR004192"
                     /db_xref="InterPro:IPR005805"
                     /db_xref="InterPro:IPR005806"
                     /db_xref="InterPro:IPR006317"
                     /db_xref="UniProt/TrEMBL:Q5H9F8"
                     /translation="SVAARSGPFAPVLSATSRGVAGALRPLVQATVPATPEQPVLDLK
                     RPFLSRESLSGQAVRRPLVASVGLNVPASVCYSHTDVKVPDFSEYRRLEVLDSTKSSR
                     ESTEARKGFSYLVTGVTTVGVAYAAKNAVTQFVSSMSASADVLALAKIEIKLSDIPEG
                     KNMAFKWRGKPLFVRHRTQKEIKQEAAVELSQLRDPQHDLDRVKKPEWVILIGVCTHL
                     GCVPIANAGDFGGYYCPCHGSHYDASGRIRLGPATLNLEVPTYEFTSDDMVIVG"
     gene            join(46614. .46832,92595. .92686,101321. .101412,
                     105557. .105676,111492. .111660,113544. .113774,
                     114913. .115035,116407. .118844)
                     /locus_tag="RP3-370M22.1-001"
     mRNA            join(46614. .46832,92595. .92686,101321. .101412,
                     105557. .105676,111492. .111660,113544. .113774,
                     114913. .115035,116407. .118844)
                     /locus_tag="RP3-370M22.1-001"
     CDS             join(92609. .92686,101321. .101412,105557. .105676,
                     111492. .111660,113544. .113774,114913. .115035,
                     116407. .116586)
                     /locus_tag="RP3-370M22.1-001"
                     /standard_name="OTTHUMP00000028837"
                     /codon_start=1
```

```
                     /protein_id="CAI42713.1"
                     /db_xref="GI:57209988"
                     /db_xref="GOA:O75791"
                     /db_xref="InterPro:IPR000980"
                     /db_xref="InterPro:IPR001452"
                     /db_xref="UniProt/Swiss-Prot:O75791"
                     /translation="MEAVAKFDFTASGEDELSFHTGDVLKILSNQEEWFKAELGSQEG
                     YVPKNFIDIQFPKWFHEGLSRHQAENLLMGKEVGFFIIRASQSSPGDFSISVRHEDDV
                     QHFKVMRDNKGNYFLWTEKFPSLNKLVDYYRTNSISRQKQIFLRDRTREDQGHRGNSL
                     DRRSQGGPHLSGAVGEEIRPSMNRKLSDHPPTLPLQQHQHQPQPPQYAPAPQQLQQPP
                     QQRYLQHHHFHQERRGGSLDINDGHCGTGLGSEMNAALMHRRHTDPVQLQAAGRVRWA
                     RALYDFEALEDDELGFHSGEVVEVLDSSNPSWWTGRLHNKLGLFPANYVAPMTR"
     gene            109527. .110407
                     /locus_tag="RP3-370M22.5-005"
                     /pseudo
     CDS             109527. .110407
                     /locus_tag="RP3-370M22.5-005"
                     /pseudo
                     /codon_start=1
     gene            join(140545. .141033,Z83847.1:21027. .31898)
                     /locus_tag="RP3-496C20.4-004"
     mRNA            join(140545. .141033,Z83847.1:21027. .21194,
                     Z83847.1:21746. .21867,Z83847.1:23149. .23822,
                     Z83847.1:31375. .31898)
                     /locus_tag="RP3-496C20.4-004"
     CDS             join(140545. .141033,Z83847.1:21027. .21194,
                     Z83847.1:21746. .21867,Z83847.1:23149. .23822,
                     Z83847.1:31375. .31424)
                     /locus_tag="RP3-496C20.4-004"
                     /standard_name="OTTHUMP00000028840"
                     /codon_start=1
                     /protein_id="CAI42715.1"
                     /db_xref="GI:57209989"
                     /db_xref="UniProt/TrEMBL:Q8NEG4"
                     /translation="MAESQLNCLDEAHVNEKVTEAQAAFYYCERRRAALEALLGGGEQ
                     AYRERLKEEQLRDFLSSPERQALRAAWSPYEDAVPAANARGKSKAKAKAPAPAPAESG
                     ESLAYWPDRSDTEVPPLDLGWTDTGFYRGVSRVTLFTHPPKDEKAPHLKQVVRQMIQQ
                     AQKVIAVVMDLFTDGDIFQDIVDAACKRRVPVYIILDEAGVKYFLEMCQDLGLTDFRI
                     RNIRVRSVTGVGFYMPMGRIKGTLSSRFLMVDGDKVATGSYRFTWSSSHVDRNLLLLL
                     TGQNVEPFDTEFRELYAISEEVDLYRQLSLAGRVGLHYSSTVARKLINPKYALVSGCR
                     HPPGEMMRWAARQQREAGGNPEGQEEGASGGESAWRLESFLKDLVTVEQVLPPVEPIP
                     LGELSQKDGRMVSHMHRDLKPKSREAPSRNGMGEAARGEAAPARRFSSRLFSRRAKRP
                     AAPNGMASSVSTETSEVEFLTGKRPNENSSADISGKTSPSSAKPSNCVIS"
ORIGIN

  Query Match           25.5%;  Score 510.4;  DB 8;  Length 143747;
  Best Local Similarity   76.2%;  Pred. No. 2e-70;
  Matches  808;  Conservative    0;  Mismatches  186;  Indels   66;  Gaps   12;

Qy        977 TACAGGCATACCTTGGAGATACT---GTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
              |||||||| |||||| |||||||    ||||| | ||||||| |   ||  ||||||  |||
Db       3144 TACAGGCACACCTTGAAGATACTGCGGTGGGCTGGGTTCCAGATTACTGCAATAAAGCAA 3203

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
              ||||  |               ||||||||  | ||||||||   ||||  | |||||||||
Db       3204 ATATTC---------------CAATAAAGCCACTCACACAAACATTTTTGTTTCCCAGTG 3248

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCT---------GTTAAGTGTGCAATAG 1144
              |||||||||||| || ||| | || |||| |||||           ||||| |||||||
Db       3249 CATATAAAAGTTATGTTTACATTAGACTGCAGTCTACTTAGTAGACTAAGTATGCAATAC 3308

Qy       1145 TGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTA----AAAAAT 1200
                |  || ||||||||        | |||||  | ||| |||||| |||    |||||
Db       3309 CATGCTGCCTAAAAAATATATATACATAATTAAATAATACTTTATTGCTAATTTAAAAAG 3368

Qy       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260
              |||||||||||| ||||| ||||| ||||  | |||||||||||||||||| |||||| ||
Db       3369 GCTAACAATCATCTGAGCCTTCAGCGAGTCTTCATCTTTTTGCTGGTGGAGGGTCTTGTC 3428

Qy       1261 TTATTGATGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTG 1309
              |   || |||  | |           ||| |   | |||||||| || ||   ||||||
Db       3429 TCCATGTTGAGGGCTGCTGATAGATCAGGGTGGTTGCTGCTGAAGGTTGGGAAAGCTGTG 3488

Qy       1310 GCAGTTTCTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCAT 1364
              ||| ||| |||||     |||||| | |||  ||||| ||||  ||||||||||||||
Db       3489 GCAATTTTTTAAAGTAAGACAACAATAAAGCCTGCAACATCAACTGACTCTTCCTTTCAC 3548

Qy       1365 GAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTC 1424
              |||||||| ||| |||  |||||||| ||||||||||||||||  ||||||||||||||
Db       3549 AAAAGATTTATCTGTAGCATGTGATGCTGTTTGATAGCATTTTACCCACAGTAGAACTTC 3608

Qy       1425 TTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATAT 1483
              ||| |||||||||| |||||||||| || |||||       || |    ||||||| || |
Db       3609 TTTCAAAATTGGAGTCAATCCTCTCTAATCCTGCCACTGGTTTATTTACTAAGTTTATGT 3668

Qy       1484 AATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAG 1543
              |||||||||||| | |||||| |||| ||||||| |||||||  |||||| |||||||||
Db       3669 AATATTCTGAATTCTTTGTTGGCATTGCAACAATGTTCACAGCTTCTTCA-CAGGAGTAG 3727

Qy       1544 ATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCT 1603
              ||| || ||||       | |    |||||||||||||||||||| | |||||||||
Db       3728 ATTTCAACTCAAGAAACCATTT----TCTTTGCTCATCCATAAGAAGTATCTCCTCATCT 3783

Qy       1604 GTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTT 1663
              |||||||||| |||||||||||||||||| |||||| |||||||      || |||||
Db       3784 GTTCAAGTTTGATCATGAGATTGCAGCAATCCAGTCACATCTTCAGG-----CTCCACTT 3838

Qy       1664 TTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCT 1723
              ||||||  |||||||||||| ||||  |||||||||  |||||||||||||  ||||||||
Db       3839 CTAATTCTAGTTCTCTTGCTATTTCCGCCACATCTGCAGTTCCTTCCTCCACTGAAGTCT 3898

Qy       1724 TGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATA 1783
              ||||||  ||| |||||||||||  |||||||||||| |||||||  |||| ||||||||||
Db       3899 TGAACCCCTCAAAGTCATCCATAAGGGTTGGAATCCACTTCTTTCAAACTCCTGTTAATG 3958

Qy       1784 TTTATATTTTG---ACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGA 1840
              ||  |  ||||||     |||||||||||||||  |||||||||||||||  || ||  ||||||
Db       3959 TTGAAATTTTGGCTTCCTCCCATGAATCACAAATGTTCTTAATGGCATCTAGAGTGGTGA 4018

Qy       1841 ATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACT 1900
              |||  |||||| |||||||| |||||||| |  |||||||||     |||||||  ||||
Db       4019 ATCTTTTCCAGAAGGTTTTTAATTTACTTTGCTCAGATCCAT-----CAGAGGAATCACT 4073

Qy       1901 TTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGA 1960
              ||  |||  ||  |||||| ||||| |||||||||||||| ||||||| ||||||||||||
Db       4074 ATCTATGGAAGATATAGCTTTATGAAATGTATTTCTTCAGTAATAAGACTTGAAAGTTGA 4133

Qy       1961 AATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||  ||||||||||||||    |||  |  ||||  |    ||||
Db       4134 AATCACTCCTTGATCCATGGGCTGAAGAATAAACACTGTG 4173
```

```
RESULT 32
AC022101/c
LOCUS       AC022101              173691 bp    DNA     linear   PRI 29-NOV-2001
DEFINITION  Homo sapiens chromosome 5 clone CTC-210G5, complete sequence.
ACCESSION   AC022101
VERSION     AC022101.4  GI:17149373
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 173691)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173691)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 173691)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (29-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 29, 2001 this sequence version replaced gi:7711682.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.2.
FEATURES             Location/Qualifiers
     source          1. .173691
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-210G5"
ORIGIN

  Query Match             25.4%;  Score 508.2;  DB 8;  Length 173691;
  Best Local Similarity   78.5%;  Pred. No. 4e-70;
  Matches  762;  Conservative    0;  Mismatches  163;  Indels   46;  Gaps   11;

Qy       1054 ACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTAT 1113
             |||||||||| |||||||||    |||||| |||| ||||||||||||  || || | |
Db     145937 ACAATAAAGTAAGTCACACAAACTTTTTGGTTTCCTAGTGCATATAAACATTATGTTCAC 145878

Qy       1114 ACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAA 1173
             |||   || ||| || ||||||||||||||   ||||||||||||   | |||| ||||
Db     145877 ACTTTGCTATAGCCTATTAAGTGTGCAATAACATTATGTCTAAAAATGTATATACTTTAA 145818

Qy       1174 TTTTAAAATGC-TTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGT 1232
             ||| ||||| | |||||| |||||||| || |||||||| || || |||| ||||||||
Db     145817 TTTAAAAATACTTTTATTTCTAAAAAATACTTACAATCATCTGGGCCTTCAATGAGTTGT 145758

Qy       1233 AATCTTTTTGCTGGTGGAAGGTCTTTTCTTAT---------TGATGACTGATC-GGGGGT 1282
             ||| ||||||||| ||||| ||||| | | ||          ||  |||||||| ||| |
Db     145757 AATATTTTTGCTGATGGAATGTCTTCTTTGATGTGGATGGCTGTGGACTGATCAGGGTGG 145698

Qy       1283 CAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCT-----TAAAACAACAGTGAAGATT 1337
               | |||| ||| || |||||||| ||||| |||||     ||||||||||||||| ||
Db     145697 TGGTTGCTAAAGATTGGGGTGGCTCTGGCAATTTCTTAAAATAAAACAACAGTGAAGTTT 145638

Qy       1338 GCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTG 1397
             ||   ||    |||||    ||||  |||||||||||||| |||| ||  ||||| ||||
Db     145637 GCTGCATAGACTGACT----CTTTTGTGAAAGATTTCTCTGTAGTATGCAATGCTGTTTG 145582

Qy       1398 ATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTG 1456
             |||||||||||  || |||||||||||||| ||||||||| |||||||| |||||||||||
Db     145581 ATAGCATTTTACCCATAGTAGAACTTCTTTCAAAATTGGAGTCAATCCTTTCAAACCCTG 145522

Qy       1457 CTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAA 1516
             ||      || |  | ||||||| || |||| |||||||||| |||||||||||||| ||
Db     145521 CTGCTGCTTTATCAACTAAGTTTATGTAATTTTCTGAATCCTTTGTTGTCATTTCAGCAC 145462

Qy       1517 TTTTCACAGTGTCT--------TCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGA 1568
             | |||||||| |||        |||||||||||||||| ||||||       | |  |
Db     145461 TGTTCACAGTATCTTCACCAGGTCACCAGGAGTAGATTCAATCTCAAGAAACCACTT--- 145405

Qy       1569 ATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCA 1628
              ||||||| |||||||||||| || |||||||||| |||||||  ||||||||||
Db     145404 -TCTTTGCCCATCCATAAGAAGCAACTCCTCATCTGTTAAAGTTTTAGTATGAGATTGCA 145346

Qy       1629 GCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTC 1688
             ||||| ||||||  | ||||||     || ||||| |||||  |||||| || || ||||
Db     145345 GCAATTCAGTCACATGTTCAGG-----CTCCACTTCTAATTAGAGTTCTTTTACTATTTC 145291

Qy       1689 TACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAG 1748
             ||| ||||||| |||  ||||||| | |||||||||||||||||| ||||||||||||||
Db     145290 TACTACATCTGCGGTAACTTCCTCTACTGAAGTCTTGAACCTCTCAAAGTCATCCATGAG 145231

Qy       1749 GGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGA 1805
             |||||||| | |||||||||||| || |||||||| | ||||| |||    ||||||||||
Db     145230 GGTTGGAACCAACTTCTTCCAAACTCTTGTTAATACTGATATTATGACCTCCTCCCATGA 145171

Qy       1806 ATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTT 1865
             |||||| |||||||||||||||||| | |||||||| ||||||| ||||||||||||||
Db     145170 ATCATGCATGTTCTTAATGGCACCTAGGATGGTGAATTCTTTCCAGAAGGTTTTCAATTT 145111

Qy       1866 ACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGG 1925
             |||| | |||||| |||     ||||  |  |||| ||  || ||| ||||||||||  |
Db     145110 ACTTTGCCCAGATTCAT-----CAGAAAAATCACTATCTGTGGCAGCTATAGCCTTACAG 145056

Qy       1926 AATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGC 1985
             |||||||||||| |||||||| |||||||| | ||||||||||||||||||    |  |
Db     145055 AATGTATTTCTTAAATAATAAGACTTGAAAGTGGTAATTACTCCTTGATCCATAGGCCAC 144996

Qy       1986 AAAATAGATGT 1996
                | |  |||||
Db     144995 AGCACTGATGT 144985


RESULT 33
HS366N23/c
LOCUS       HS366N23               98517 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP3-366N23 on chromosome 6q27
            Contains the UNC93A gene for unc-93 homolog A (C. elegans), the
            gene for NYD-TSPG protein,  a Mitotic Feedback Control Protein
            MADP2 Homolog LIKE pseudogene and the 3' end  of the TCP10 gene for
            t-complex 10 (mouse), complete sequence.
ACCESSION   AL021331
VERSION     AL021331.1  GI:3355532
KEYWORDS    HTG; C46F11.1; MADP2; Mitotic Feedback Control; NYD-TPSG;
            t-complex; TCP10; UNC-93.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 98517)
  AUTHORS   Phillips,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 30, 1998 this sequence version replaced gi:3334553.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP3-366N23 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            ----------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .98517
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704N23366"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q27"
                     /clone="RP3-366N23"
                     /clone_lib="RPCI-3"
     gene            29762. .54362
                     /gene="UNC93A"
                     /locus_tag="RP3-366N23.1-001"
     mRNA            join(29762. .29925,32866. .33047,34381. .34610,36294. .36419,
                     42283. .42482,44264. .44399,46128. .46259,53536. .54362)
                     /gene="UNC93A"
                     /locus_tag="RP3-366N23.1-001"
                     /product="unc-93 homolog A (C. elegans)"
                     /note="match: ESTs: T60316 T60341 W78931
                     match: cDNAs: AJ508812"
     CDS             join(29839. .29925,32866. .33047,34381. .34610,36294. .36419,
                     42283. .42482,44264. .44399,46128. .46259,53536. .53801)
                     /gene="UNC93A"
                     /locus_tag="RP3-366N23.1-001"
                     /standard_name="OTTHUMP00000017623"
                     /note="match: proteins: CAD48541"
                     /codon_start=1
                     /product="unc-93 homolog A (C. elegans)"
                     /protein_id="CAI42706.1"
                     /db_xref="GI:57208158"
                     /db_xref="UniProt/TrEMBL:Q5JZD6"
                     /translation="MDRSLRNVLVVSFGFLLLFTAYGGLQSLQSSLYSEEGLGVTALS
                     TLYGGMLLSSMFLPPLLIERLGCKGTIILSMCGYVAFSVGNFFASWYTLIPTSILLGL
                     GAAPLWSAQCTYLTITGNTHAEKAGKRGKDMVNQYFGIFFLIFQSSGVWGNLISSLVF
                     GQTPSQETLPEEQLTSCGASDCLMATTTTNSTQRPSQQLVYTLLGIYTAVLMIAAFLQ
                     PIRDVQRESEGEKKSVPFWSTLLSTFKLYRDKRLCLLILLPLYSGLQQGFLSSEYTRS
                     YVTCTLGIQFVGYVMICFSATDALCSVLYGKVSQYTGRAVLYVLGAVTHVSCMIALLL
                     WRPRADHLAVFFVFSGLWGVADAVWQTQNNALYGVLFEKSKEAAFANYRLWEALGFVI
                     AFGYSMFLCVHVKLYILLGVLSLTMVAYGLVECVESKNPIRPHAPGQVNQAEDEEIQT
                     KM"
     gene            63467. .80236
                     /gene="RP3-366N23.3"
                     /locus_tag="RP3-366N23.3-001"
     mRNA            join(63467. .63569,76996. .77152,78454. .80236)
                     /gene="RP3-366N23.3"
                     /locus_tag="RP3-366N23.3-001"
                     /note="match: ESTs: BG720252 BI462156 BI463431 BI560373
                     match: cDNAs: AY026506"
     CDS             join(63523. .63569,76996. .77152,78454. .80028)
                     /gene="RP3-366N23.3"
                     /locus_tag="RP3-366N23.3-001"
                     /standard_name="OTTHUMP00000017625"
                     /note="match: proteins: Q9BWV7"
                     /codon_start=1
                     /protein_id="CAD92564.1"
                     /db_xref="GI:31652294"
                     /db_xref="Genew:21211"
                     /db_xref="GOA:Q9BWV7"
                     /db_xref="UniProt/Swiss-Prot:Q9BWV7"
                     /translation="MRGRDLCSSTQSQALGSLRTTTPAFTLNIPSEANHTEQPPAGLG
                     ARLQEAGVSIPPRRGRPTPTLEKKKKPHLMAEDEPSGALLKPLVFRVDETTPAVVQSV
                     LLERGWNKFDKQEQNAEDWNLYWRTSSFRMTEHNSVKPWQQLNHHPGTTKLTRKDCLA
                     KHLKHMRRMYGTSLYQFIPLTFVMPNDYTKFVAEYFQERQMLGTKHSYWICKPAELSR
                     GRGILIFSDFKDFIFDDMYIVQKYISNPLLIGRYKCDLRIYVCVTGFKPLTIYVYQEG
                     LVRFATEKFDLSNLQNNYAHLTNSSINKSGASYEKIKEVIGHGCKWTLSRFFSYLRSW
                     DVDDLLLWKKIHRMVILTILAIAPSVPFAANCFELFGFDILIDDNLKPWLLEVNYSPA
                     LTLDCSTDVLVKRKLVHDIIDLIYLNGLRNEGREASNATHGNSNIDAAKSDRGGLDAP
                     DCLPYDSLSFTSRMYNEDDSVVEKAVSVRPEAAPASQLEGEMSGQDFHLSTREMPQSK
                     PKLRSRHTPHKTLMPYASLFQSHSCKTKTSPCVLSDRGKAPDPQAGNFVLVFPFNEAT
                     LGASRNGLNVKRIIQELQKLMNKQHS"
     polyA_signal    80171. .80176
                     /gene="RP3-366N23.3"
                     /locus_tag="RP3-366N23.3-001"
     polyA_site      80236
                     /gene="RP3-366N23.3"
                     /locus_tag="RP3-366N23.3-001"
     gene            85821. .86402
                     /locus_tag="RP3-366N23.4-001"
                     /pseudo
     CDS             85821. .86402
                     /locus_tag="RP3-366N23.4-001"
                     /note="match: proteins: P79935 Q13257"
                     /pseudo
                     /codon_start=1
     polyA_site      complement(95380)
     gene            join(complement(AL592444.18:16183. .26342),
                     complement(95382. .97761))
                     /gene="TCP10"
                     /locus_tag="RP3-366N23.5-001"
     mRNA            join(complement(AL592444.18:26213. .26342),
                     complement(AL592444.18:24643. .24787),
                     complement(AL592444.18:22953. .23168),
                     complement(AL592444.18:19787. .19924),
                     complement(AL592444.18:18378. .18536),
                     complement(AL592444.18:17798. .17909),
                     complement(AL592444.18:16183. .16283),
                     complement(97644. .97761),complement(95382. .96645))
                     /gene="TCP10"
                     /locus_tag="RP3-366N23.5-001"
                     /product="t-complex 10 (mouse)"
                     /note="match: ESTs: AA431681.1 BI828613.1
                     match: cDNAs: U03399 U03399.1"
     polyA_signal    complement(95407. .95412)
                     /gene="TCP10"
                     /locus_tag="RP3-366N23.5-001"
     CDS             join(complement(AL592444.18:24643. .24705),
                     complement(AL592444.18:22953. .23168),
                     complement(AL592444.18:19787. .19924),
                     complement(AL592444.18:18378. .18536),
                     complement(AL592444.18:17798. .17909),
                     complement(AL592444.18:16183. .16283),
                     complement(97654. .97761))
                     /gene="TCP10"
                     /locus_tag="RP3-366N23.5-001"
                     /standard_name="OTTHUMP00000017628"
                     /note="match: proteins: Q9NQ19"
                     /codon_start=1
                     /product="t-complex 10 (mouse)"
                     /protein_id="CAI42705.1"
                     /db_xref="GI:57208157"
                     /db_xref="UniProt/TrEMBL:Q5JR60"
                     /translation="MEKTPAAAEVPREDSNAGEMPSLQQQITSLHQELGRQQSLWADI
                     HRKLQSHMDALRKQNRELREELRGLQRQQWEAGKKPAASPHAGRESHTLALEPAFGKI
                     SHLSADEDTTPKYAGRKSQSATLLGQRWSSNHLAPPKPMSLKTERINSGKTPPQEDRE
                     KSPPGRRQDRSPAPTGRPTPGAERRGVSEDGKIMHPSSRSPQNSGGRKSPVQASQATT
                     LQEQTAAARGADRSSSVLGSSEGGFLSRVQADEFASSAPDSAERQKNSTPMAPKKSCF
                     PTGRWNISRMDRKRPYFLMGQS"
ORIGIN

  Query Match            25.4%;  Score 507.6;  DB 8;  Length 98517;
  Best Local Similarity  75.6%;  Pred. No. 6.2e-70;
  Matches  791;  Conservative    0;  Mismatches  204;  Indels   51;  Gaps   11;

Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
              ||||||||  |||||| ||||||||||  ||||||||||  ||||||||||
Db      92618 TACAGGCAGGCCTTGGGGATACTGTGGATTTGGTTCCAGGCCACCACAATA--------- 92568

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096
                 |||   ||||  ||| | ||  ||||||||| | |||||  | ||||||| |
Db      92567 ------AAGCCCATATTGCAACAGAGCAAGTCACACAAATATTTTGTTTCCCCAGTGCCT 92514

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156
              |||||||||| | ||||| || |||||||||| ||||||||  |||||  ||||||||||
Db      92513 ATAAAAGTTTGGTTTATATTATACTGTAGTCTATTAAGTGTATAATAGCATTATGTCTAA 92454

Qy       1157 AA-AAACACATACCTTAATTTTAAAATGCTTTATT--ACTAAAAAATGCTAACAATCATT 1213
              || |  ||||| |||||||| |||||  | ||||    ||||||||||||  |||||
Db      92453 AATATGAACATATCTTAATTTAAAAATAGTGTATTGATTTAAAAAATGCTAATGATCATC 92394

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTG 1273
              |||||  |||||||||| ||||  |||||||||||||| || |||  ||   || |||  |
```

```
Db      92393 TGAGCCCTCAGTGAGTTATAATTGTTTTGCTGGTGGAGGGCCTTGCCTGGATGTTGATGG 92334

Qy       1274 AT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCT---- 1318

Db      92333 CTGCTGACTGATCAGGGTGGAGGTTGCTGAAGGTTGGGGTGACTGTGGCAATATCTGAAA 92274

Qy       1319 TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTC 1378

Db      92273 TAAGACCACAATGAAATTTACTGCATTGACTGACTCTTCCTCTCATGAAAGGTTTCTCTG 92214

Qy       1379 TAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA- 1437

Db      92213 TAGCATGCGATGCTGTTTGATAGCATTTTACCCACAGTACAACTTCTTTTAAAATTGGAG 92154

Qy       1438 TCAATCCTCTCAAACCCTGCTC-TGCTTTAACAACCTAAGTTAATATAATATTCTGAATC 1496

Db      92153 TCAATCCCCTCAAACCCTACCCACTGCCCCATCAACTAAGTTCATGGAATATTCTAAATC 92094

Qy       1497 CATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATT 1556

Db      92093 ATTTGTTGTCATATCAATAACGTTCACAGCATCTTCACCAGGAGTCAATTCTATCTCAAG 92034

Qy       1557 TCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTAT 1616

Db      92033 AAACCACTT----TCTTTCCTCAACCATAAGAAGCCACTCCTGATCTGTTCAAATTTTAT 91978

Qy       1617 CATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTC 1676

Db      91977 CCTGAGATTGCAGCAATTCAGTCACATTTTTAGGCCCCACTTC-----TAATTCTAGTTC 91923

Qy       1677 TCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAA 1736

Db      91922 TTTTGTTATTTCCACCTCATCTACAGTTACTTTCTGCAGTGAAGGCTTGAACCCCTCAAA 91863

Qy       1737 GTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATT--TTG 1794

Db      91862 ATCATTCATGAGGAATGAAATCAACTTCTTCTAAACTCCTGTTAATGTTGATATTGGTTC 91803

Qy       1795 ACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAG 1854

Db      91802 TCCTCCCGTGAATCATGAATGTTGTTAATGGCATCTAGAATGGTGAATCCTTTCTAGATG 91743

Qy       1855 GTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTA 1914

Db      91742 GTTTTTAATTGACTTTGCCCAGATCCAT-----CTGAGGAATCACTATCTATGGCAGCTA 91688

Qy       1915 TAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGAT 1974

Db      91687 TAGCCTTTTGAGATGTATTTCTTAAGTAACAAGACTTGAAAGTTGAAATTACTCCTTGAT 91628

Qy       1975 CCATTTTCTGCAAAATAGATGTTGTG 2000

Db      91627 CCATGGGCTGCAGAATGGATGTGGTG 91602


RESULT 34
AC084703/c
LOCUS       AC084703              110907 bp    DNA     linear   PRI 28-MAR-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-1152D21, complete sequence.
ACCESSION   AC084703
VERSION     AC084703.7  GI:19774531
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 110907)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-1152D21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 110907)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 110907)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 110907)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 28, 2002 this sequence version replaced gi:18698799.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10865
                Center clone name: 1152_D_21
            --------------

            Only the first 110.9 kb of this clone are being submitted.
            The remainder overlaps accession number AC090868 [WICGR project
            L13154].
FEATURES                Location/Qualifiers
     source             1. .110907
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="15"
                        /map="15"
                        /clone="RP11-1152D21"
                        /clone_lib="RPCI-11 Human Male BAC"
     repeat_region      143. .166
                        /rpt_family="AT_rich"
     repeat_region      complement(167. .489)
                        /rpt_family="MER2"
     repeat_region      490. .512
                        /rpt_family="AT_rich"
     repeat_region      complement(776. .985)
                        /rpt_family="L2"
     repeat_region      1051. .7210
                        /rpt_family="L1PA3"
     repeat_region      7986. .8009
                        /rpt_family="AT_rich"
     repeat_region      complement(8312. .8382)
                        /rpt_family="MER105"
     repeat_region      complement(8383. .8530)
                        /rpt_family="MER5A"
     repeat_region      complement(8531. .8578)
                        /rpt_family="MER105"
     repeat_region      8632. .8660
                        /rpt_family="AT_rich"
     repeat_region      9055. .9211
                        /rpt_family="FAM"
     repeat_region      9921. .9956
                        /rpt_family="(A)n"
     repeat_region      complement(10326. .10612)
                        /rpt_family="L1PA7"
     repeat_region      complement(10704. .10749)
                        /rpt_family="MIR"
     repeat_region      complement(10818. .10922)
                        /rpt_family="FLAM_A"
     repeat_region      11092. .11128
                        /rpt_family="AT_rich"
     repeat_region      11315. .11565
                        /rpt_family="Charlie7"
     repeat_region      12263. .12512
                        /rpt_family="L1MA9"
     repeat_region      12824. .12979
                        /rpt_family="L1M4c"
     repeat_region      complement(13134. .13244)
                        /rpt_family="L1ME"
     repeat_region      13271. .13680
                        /rpt_family="MLT1I"
     repeat_region      13782. .13804
                        /rpt_family="AT_rich"
     repeat_region      complement(13824. .13955)
                        /rpt_family="MIR3"
     repeat_region      13986. .14025
                        /rpt_family="AT_rich"
     repeat_region      complement(14176. .14470)
                        /rpt_family="AluSx"
     repeat_region      15104. .15130
                        /rpt_family="AT_rich"
     repeat_region      complement(15557. .15615)
                        /rpt_family="MIR"
     repeat_region      complement(15651. .15993)
                        /rpt_family="MLT1I"
     repeat_region      complement(16131. .16421)
                        /rpt_family="AluJb"
     repeat_region      complement(16486. .16993)
                        /rpt_family="L3"
     repeat_region      complement(17003. .17213)
                        /rpt_family="MER58C"
     repeat_region      complement(17214. .17676)
                        /rpt_family="L3"
     repeat_region      complement(17677. .17985)
                        /rpt_family="AluSc"
     repeat_region      complement(17986. .18090)
                        /rpt_family="L3"
     repeat_region      18091. .23830
                        /rpt_family="L1PA7"
     repeat_region      23834. .23856
                        /rpt_family="(TAAA)n"
     repeat_region      23975. .24193
                        /rpt_family="MER58A"
     repeat_region      24352. .24675
                        /rpt_family="L2"
     repeat_region      complement(25054. .25564)
                        /rpt_family="MER119"
     repeat_region      complement(25710. .25886)
                        /rpt_family="MER58A"
     repeat_region      26048. .26488
                        /rpt_family="L1MC4"
     repeat_region      27129. .27235
                        /rpt_family="MER103"

 Query Match             25.3%;  Score 506.2;  DB 8;  Length 110907;
 Best Local Similarity   75.3%;  Pred. No. 9.7e-70;
 Matches  793;  Conservative    0;  Mismatches  208;  Indels   52;  Gaps   11;

Qy        974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
              || | |||| ||||||| ||||||||||||||  |||||| | ||||||||||| |||
Db      62627 TAATGCAGGTATACCTTAGAGATACTGTGGGTTCAGTTCCAGAGCACCACAATAAAGCAA 62568

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
              || |          |||    |       | ||||| ||| |   |||| | ||||||||
Db      62567 ATGTA--------GGAATAAATGTAGGAATGAGTTACATGAAGTTTTTAGTTTCCCAGTA 62516

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
              |||||||||||| || ||| |||| |||||||||| |||| ||||| |||| |||||||
Db      62515 CATATAAAAGTTATGTTTACACTATACTGTAGTCTATTAAATGTGCTATAGCATTATGTC 62456

Qy       1154 TAAAAAAAC-----ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAA 1208
              |||||||||     |||||||||||||  ||||| ||||| || |||||||||||||  |
Db      62455 TAAAAAAACTATGTACATACCTTAATTCAAAAATACTTTAGTAATAAAAAATGCTAATGA 62396

Qy       1209 TCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAG--------------GT 1254
              |||| ||| | |||||  |||||||||||||||||||||||| |              |
Db      62395 TCATCTGATCTTTCAGCAAGTTGTAATCTTTTTGCTGGTGGAGGGAGGGTCTTGCCGTGA 62336

Qy       1255 CTTTTCTTATTGATGACTGATCGGGGGTCAGG-TGCTGAAGCTTAGGGTGGCTGTGGCAG 1313
                ||  |   || |||| ||||  ||    || |||||||| || || ||||||||||
Db      62335 TGTTAATGGCTGCTGACAGATCAAGGTGATGGTTGCTGAAGGTTGGGATGGCTGTGGCGT 62276

Qy       1314 TTTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTT 1373
                | | |||| | ||||  ||||   |    |||||| |||| ||||||||||||||||
Db      62275 AAAACTCAGACAAAAATGAAATTTGCCTCA----TTGACCCTTCTTTTCATGAAAGATTT 62220

Qy       1374 CTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAAT 1433
              |||| |||  || ||| || ||||||||||||||   ||||||  |||||||| | |||
```

```
Db      62219 CTCTGTAGCATGAGATACTGTTTGATAGCATTTTACTTACAGTAAGACTTCTTTCAGAAT 62160

Qy       1434 TG-GATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTG 1492
              ||  ||| ||||||||||||||||||      || |  | |||||||||| | |||||||
Db      62159 TGAAATCCGTCCTCTCAAACCCTGCTGGTGCTTTATCAACTAAGTTAATGTGATATTCTA 62100

Qy       1493 AATCCATTGTTGTCATTTCAACAATTTTCACAG--TGTCTTCACCAGGAGTAGATTCCAT 1550
              | |   |||||||||||||| |||| |||| ||  | | |||||||||| || ||||||
Db      62099 AGTTGTTTGTTGTCATTTCAGCAATGTTCATAGCATCTTTTCACCAGGAGCAGTTTCCAT 62040

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              || |       |  |    ||||| ||||||| |||||||  | ||||||||| ||||||
Db      62039 CTTAAGAAACTATTT----TCTTTCCTCATCCGTAAGAAGTGACTCCTCATCTATTCAAG 61984

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              |||  |||||||||||||| || ||||||  |||| ||||     | |||||||| |||
Db      61983 TTTGGTCATGAGATTGCAGCCATTCAGTCACATCTTTAGGC-----TCCACTTTTAGTTC 61929

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
              ||||||  ||||  |||||||||||||||  |||||||||| || ||||||||||||||
Db      61928 TAGTTCTTTTGCCCTTTCTACCACATCTGCAGTTCCTTCCTTCACTGAAGTCTTGAACCC 61869

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              ||| ||||||||||||||| |||| ||| ||||| |||||| |||||||||| || ||||
Db      61868 CTCAAAGTCATCCATGAGGTTTGGGATCAACTTCCTCCAAACTCCTGTTAATGTTGATAT 61809

Qy       1791 TTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
              |||||   |||||||||||||||||||||||||| ||| || | ||||| ||| ||||
Db      61808 TTTGACCTCCTCCCATGAATCATGAATGTTCTTAATAGCATCTAGGATGGTCAATTCTTT 61749

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              | | || ||||||| | ||||| | ||||||||        ||||||  |||| |  | |
Db      61748 CTAGAAAGTTTTCAGTCTACTTTGCCCAGATCCA-----GTAGAGGAATCACTATTTAAG 61694

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
              ||| || ||||||||| ||||||| |||| ||||||||  |||||||||||| ||||||
Db      61693 GCAGCTAAAGCCTTATGAAATGTATATCTTAAATAATAATCCTTGAAAGTTGATATTACT 61634

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||| |||||   ||||| ||| |||||||||
Db      61633 CCTTGCTCCATGGACTGCAGAATGGATGTTGTG 61601
```

```
RESULT 35
AC013752/c
LOCUS       AC013752               183794 bp    DNA     linear   PRI 29-DEC-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-396J17, complete sequence.
ACCESSION   AC013752
VERSION     AC013752.10  GI:27413935
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 183794)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-396J17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183794)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 183794)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 183794)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 29, 2002 this sequence version replaced gi:21622793.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2432
                Center clone name: 396_J_17
            --------------.
FEATURES             Location/Qualifiers
     source          1. .183794
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-396J17"
```

```
                /clone_lib="RPCI-11 Human Male BAC"
repeat_region   48. .160
                /rpt_family="MIR"
repeat_region   complement(304. .572)
                /rpt_family="LTR16C"
repeat_region   1091. .1186
                /rpt_family="MIR"
repeat_region   complement(1197. .1495)
                /rpt_family="AluJo"
repeat_region   1497. .1651
                /rpt_family="MIR"
repeat_region   complement(1653. .1838)
                /rpt_family="L2"
repeat_region   1839. .2257
                /rpt_family="MSTB"
repeat_region   complement(2258. .2315)
                /rpt_family="L2"
repeat_region   complement(2397. .2698)
                /rpt_family="L2"
repeat_region   complement(3400. .3588)
                /rpt_family="L1PA16"
repeat_region   complement(3590. .5413)
                /rpt_family="L1PA16"
repeat_region   5414. .5517
                /rpt_family="(TTTC)n"
repeat_region   complement(5518. .5804)
                /rpt_family="L1PA16"
repeat_region   complement(7391. .7969)
                /rpt_family="MER68"
repeat_region   8172. .8322
                /rpt_family="FRAM"
repeat_region   8326. .8378
                /rpt_family="(TAAA)n"
repeat_region   9897. .10803
                /rpt_family="L2"
repeat_region   11116. .11227
                /rpt_family="MIR"
repeat_region   11419. .11716
                /rpt_family="AluJo"
repeat_region   12056. .12212
                /rpt_family="FAM"
repeat_region   12365. .12506
                /rpt_family="L2"
repeat_region   complement(13509. .13604)
                /rpt_family="MIR3"
repeat_region   14272. .14440
                /rpt_family="MER117"
repeat_region   complement(15242. .15443)
                /rpt_family="L1MEc"
repeat_region   15698. .15875
                /rpt_family="LTR47A"
repeat_region   complement(15878. .16264)
                /rpt_family="MSTD"
repeat_region   complement(16265. .16482)
                /rpt_family="L1MEc"
repeat_region   complement(16483. .16787)
                /rpt_family="AluJo"
repeat_region   complement(16788. .17182)
                /rpt_family="L1MEc"
repeat_region   complement(17237. .17283)
                /rpt_family="L1MEc"
repeat_region   complement(17284. .17593)
                /rpt_family="AluJb"
repeat_region   17601. .17626
                /rpt_family="(TG)n"
repeat_region   complement(17627. .18455)
                /rpt_family="L1MEc"
repeat_region   19065. .19124
                /rpt_family="(TATATG)n"
repeat_region   complement(19217. .19428)
                /rpt_family="MER106A"
repeat_region   19724. .19779
                /rpt_family="AT_rich"
repeat_region   19918. .20460
                /rpt_family="LTR9"
repeat_region   complement(20465. .20783)
                /rpt_family="L1M4"
repeat_region   complement(20784. .21266)
                /rpt_family="LTR40a"
repeat_region   complement(21267. .21353)
                /rpt_family="L1M4"
repeat_region   21647. .21696
                /rpt_family="(CA)n"
repeat_region   complement(21949. .22126)
                /rpt_family="MER5B"
repeat_region   23039. .23392
                /rpt_family="MLT1A1"
repeat_region   complement(23809. .24025)
                /rpt_family="L1MC4"
repeat_region   complement(24161. .24519)
                /rpt_family="THE1C"
repeat_region   complement(26065. .26217)
                /rpt_family="MER5A"
repeat_region   complement(26239. .26334)
                /rpt_family="L1MA5"
repeat_region   26705. .26733
                /rpt_family="AT_rich"
repeat_region   26775. .26823
                /rpt_family="MER111"

 Query Match            25.3%;  Score 506.2;  DB 8;  Length 183794;
 Best Local Similarity  75.3%;  Pred. No. 8.1e-70;
 Matches  793;  Conservative    0;  Mismatches  208;  Indels   52;  Gaps   11;

Qy       974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
             || | |||| ||||||| |||||||||||||||  |||||| | ||||||||||| |||
Db    163619 TAATGCAGGTATACCTTAGAGATACTGTGGGTTCAGTTCCAGAGCACCACAATAAAGCAA 163560

Qy      1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
             || |          |||    |      | ||||| |||  |   |||| | ||||||||
Db    163559 ATGTA--------GGAATAAATGTAGGAATGAGTTACATGAAGTTTTTAGTTTCCCAGTA 163508

Qy      1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
             |||||||||||| || ||| |||| |||||||||| |||| ||||| |||| |||||||
Db    163507 CATATAAAAGTTATGTTTACACTATACTGTAGTCTATTAAATGTGCTATAGCATTATGTC 163448

Qy      1154 TAAAAAAAC-----ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAA 1208
             |||||||||     |||||||||||||  ||||| ||||| || |||||||||||||  |
Db    163447 TAAAAAAACTATGTACATACCTTAATTCAAAAATACTTTAGTAATAAAAAATGCTAATGA 163388

Qy      1209 TCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAG--------------GT 1254
             |||| ||| | |||||  |||||||||||||||||||||||| |              |
Db    163387 TCATCTGATCTTTCAGCAAGTTGTAATCTTTTTGCTGGTGGAGGGAGGGTCTTGCCGTGA 163328

Qy      1255 CTTTTCTTATTGATGACTGATCGGGGGTCAGG-TGCTGAAGCTTAGGGTGGCTGTGGCAG 1313
               ||  |   || |||| ||||  ||    || |||||||| || || ||||||||||
Db    163327 TGTTAATGGCTGCTGACAGATCAAGGTGATGGTTGCTGAAGGTTGGGATGGCTGTGGCGT 163268

Qy      1314 TTTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTT 1373
                | | |||| | ||||  ||||    |    |||||| |||| ||||||||||||||||
Db    163267 AAAACTCAGACAAAAATGAAATTTGCCTCA----TTGACCCTTCTTTTCATGAAAGATTT 163212

Qy      1374 CTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAAT 1433
             |||| |||  || ||| || ||||||||||||||    ||||||  |||||||| | |||
Db    163211 CTCTGTAGCATGAGATACTGTTTGATAGCATTTTACTTACAGTAAGACTTCTTTCAGAAT 163152

Qy      1434 TG-GATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTG 1492
             ||  |||  ||||||||||||||||||     || |  | |||||||||| | |||||||
Db    163151 TGAAATCCGTCCTCTCAAACCCTGCTGGTGCTTTATCAACTAAGTTAATGTGATATTCTA 163092

Qy      1493 AATCCATTGTTGTCATTTCAACAATTTTCACAG--TGTCTTCACCAGGAGTAGATTCCAT 1550
             | |   |||||||||||||| |||| |||| ||  | | |||||||||| || ||||||
Db    163091 AGTTGTTTGTTGTCATTTCAGCAATGTTCATAGCATCTTTTCACCAGGAGCAGTTTCCAT 163032

Qy      1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
             || |       |  |    ||||| ||||||| |||||||  | ||||||||| ||||||
```

```
Db      163031 CTTAAGAAACTATTT----TCTTTCCTCATCCGTAAGAAGTGACTCCTCATCTATTCAAG 162976

Qy        1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
               |||  |||||||||||||||| || ||||||  |||| ||||     | ||||||||  |||
Db      162975 TTTGGTCATGAGATTGCAGCCATTCAGTCACATCTTTAGGC-----TCCACTTTTAGTTC 162921

Qy        1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
               ||||||  ||||  ||||||||||||||||  |||||||||| || ||||||||||||||
Db      162920 TAGTTCTTTTGCCCTTTCTACCACATCTGCAGTTCCTTCCTTCACTGAAGTCTTGAACCC 162861

Qy        1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
               ||| ||||||||||||||| |||| ||| ||||| |||||| |||||||||| || ||||
Db      162860 CTCAAAGTCATCCATGAGGTTTGGGATCAACTTCCTCCAAACTCCTGTTAATGTTGATAT 162801

Qy        1791 TTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
               |||||   |||||||||||||||||||||||||| ||| || | ||||| ||| ||||
Db      162800 TTTGACCTCCTCCCATGAATCATGAATGTTCTTAATAGCATCTAGGATGGTCAATTCTTT 162741

Qy        1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
               | | || ||||||| | ||||| | |||||||||       ||||||  |||| | | |
Db      162740 CTAGAAAGTTTTCAGTCTACTTTGCCCAGATCCA-----GTAGAGGAATCACTATTTAAG 162686

Qy        1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
               |||  || |||||||| |||||||  |||| ||||||||  |||||||||||| ||||||
Db      162685 GCAGCTAAAGCCTTATGAAATGTATATCTTAAATAATAATCCTTGAAAGTTGATATTACT 162626

Qy        1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
               ||||| |||||   ||||| ||| |||||||||
Db      162625 CCTTGCTCCATGGACTGCAGAATGGATGTTGTG 162593


RESULT 36
AC022028/c
LOCUS       AC022028               156320 bp    DNA     linear   PRI 22-MAY-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-93H12, complete sequence.
ACCESSION   AC022028
VERSION     AC022028.7  GI:21070552
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 156320)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156320)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 156320)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JAN-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   4  (bases 1 to 156320)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On May 22, 2002 this sequence version replaced gi:18201761.
FEATURES             Location/Qualifiers
     source          1. .156320
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-93H12"
                     /clone_lib="RPCI-11"
ORIGIN
```

```
  Query Match           25.3%;  Score 505.2;  DB 8;  Length 156320;
  Best Local Similarity   75.5%;  Pred. No. 1.2e-69;
  Matches  743;  Conservative    0;  Mismatches  208;  Indels   33;  Gaps    8;

Qy        1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
               || |  | ||||| |||| |||||||   |||| ||  | | |||||| || ||||||
Db      112548 ATCTCAATAAAGTGAATATTGCAATAAATAAAGTCGCAGGAATTTTTTGGTTTTCCAGTG 112489

Qy        1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
               |||||||||||| || ||||| || |||||||||| |||||| | |||||| |||||||
Db      112488 CATATAAAAGTTATGTTTATATTATACTGTAGTCTATTAAGTATTCAATAGCATTATGTC 112429

Qy        1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
               |||||||         ||| | ||| |||      |  | || |||| | |||||   ||
Db      112428 TAAAAAATATACATACTTTAATTTTAAAGTACCTTATGCTCAAAAGTACTAACCCCCACC 112369

Qy        1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT---------TTTCTTAT 1264
               ||||| |||||||||||||||||||||||||||||||||| ||| |         ||  |
Db      112368 TGAGCCTTCAGTGAGTTGTAATCTTTTTGCTGGTGGAGGGTTTGCCTCAAAGTTGATGTC 112309

Qy        1265 TGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-- 1322
               |  |||||||||  | |       |||| ||| |  | ||||||||||||||||||||||
Db      112308 TCCTGACTGATCAGAGTGGTATTGCTAAAGGTAGGAGTGGCTGTGGCAGTTTCTTAAAAT 112249

Qy        1323 ---ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
                  ||||||||||  |||| | |||| |||||||||| |||||||||||| |||||| |
Db      112248 GAGACAACAGTGAAACTTGCCACATCAATTGACTCTTCTTTTCATGAAAGAATTCTCTGT 112189

Qy        1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-T 1438
               ||  ||||||| | |||| || |||||||  |||||||||||||| ||||  ||||| |
Db      112188 AGCATGTGATGTTGTTTGGTACCATTTTACTCACAGTAGAACTTCTGTGAACGTTGGAGT 112129

Qy        1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
               ||  | |||||||| |||       || |  | |||  |  ||  |||||||| || |
Db      112128 CACACTTCTCAAACCTTGCCACTTCTTCATCAACTATATGTATGGAATATTCTAAACCTT 112069

Qy        1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
               |||||||||||||||||| |||| ||| | |||||||||||| |||| ||||||
Db      112068 TTGTTGTCATTTCAACAATGTTCATAGTATTTTCACCAGGAGTATATTCTATCTCAAGAA 112009

Qy        1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
                 |  |    |||||||||||  ||||||  || |  ||| |  || |||||  ||||
Db      112008 GCCACTT----TCTTTGCTCATCTGTAAGAAACAAATGTTCACCCATTTAAGTTCGATCA 111953

Qy        1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
               ||| |||||||||| |||||    |||||||||      || |  || |||||| |||||||
Db      111952 TGACATTGCAGCAATTCAGTCTCATCTTCAGG-----CTCCCTTTCTAATTCTAGTTCTC 111898

Qy        1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
               ||||| |||||||||||| ||  ||| |||||||| ||||||||||||||||||| ||||
Db      111897 TTGCTATTTCTACCACATATGCAGTTACTTCCTCCACTGAAGTCTTGAACCTCTCAAAGT 111838

Qy        1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
               ||||||| |||||||||||| | ||||| | || ||||||| || || ||||||||||||
Db      111837 CATCCATCAGGGTTGGAATCAATTTCTTGCCAACTCCTGTTTATGTTGATATTTTGACCT 111778

Qy        1799 CC---CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855
               ||   |||||||  |  | |||| |||||||||| ||||||||||||||| || ||||
Db      111777 CCTCTCATGAATTACAAGTATTCTGAATGGCACCTAGAATGGTGAATCCTTTTCAGAAGG 111718

Qy        1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915
                ||||||||||||  | |||||||||      |||| ||  |||| || ||| ||| ||
Db      111717 -TTTCAATTTACTTTGCCCAGATCCAT-----CAGAAGAATCACTATCTATGGCAGCTAC 111664

Qy        1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
               ||||||||| | |||||||||| ||| ||||  |||||||||  ||||||  ||||||||
Db      111663 AGCCTTATGAAGTGTATTTCTTAAATGATAAAACTTGAAAGTCAAAATTATACCTTGATC 111604

Qy        1976 CATTTTCTGCAAAATAGATGTTGT 1999
               |||   ||||| ||| ||||||||
Db      111603 CATGGGCTGCAGAATGGATGTTGT 111580
```

```
RESULT 37
AL139142/c
LOCUS       AL139142              159771 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP3-436N22 on chromosome 1q24.1-25.3
            Contains the PRDX6 gene for peroxiredoxin 6, gene MGC43026, a novel
            gene, the 3' end of a novel gene (FLJ45235) and two CpG islands,
            complete sequence.
ACCESSION   AL139142
VERSION     AL139142.10  GI:15620584
KEYWORDS    HTG; FLJ45235; MGC43026; PRDX6.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 159771)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 14, 2001 this sequence version replaced gi:15590740.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP3-436N22 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .159771
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704N22436"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q24.1-25.3"
                     /clone="RP3-436N22"
                     /clone_lib="RPCI-3"
     misc_feature    1
                     /note="Clone_left_end: RP3-436N22"
     gene            join(3133. .3359,7193. .7349,11228. .11374,12122. .12268,
                     13602. .14674)
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-001"
     mRNA            join(3133. .3359,7193. .7349,11228. .11374,12122. .12268,
                     13602. .14674)
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-001"
                     /product="peroxiredoxin 6"
                     /note="match: cDNAs: Em:BC008738.1 Em:BC017252.1
                     Em:BC025421.2 Em:BC035857.1 Em:BC053550.1 Em:D14662.1"
     gene            join(3197. .3359,7193. .7838)
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-002"
     mRNA            join(3197. .3359,7193. .7838)
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-002"
                     /product="peroxiredoxin 6"
                     /note="match: ESTs: Em:CB988834.1"
     CDS             join(3265. .3359,7193. .7349,11228. .11374,12122. .12268,
                     13602. .13730)
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-001"
                     /standard_name="OTTHUMP00000032693"
                     /note="match: proteins: Sw:O08709 Sw:O35244 Sw:O77834
                     Tr:AAP21829 Tr:Q7SYQ4 Tr:Q7SYT1"
                     /codon_start=1
                     /product="peroxiredoxin 6"
                     /protein_id="CAI20936.1"
                     /db_xref="GI:56204402"
                     /translation="MPGGLLLGDVAPNFEANTTVGRIRFHDFLGDSWGILFSHPRDFT
                     PVCTTELGRAAKLAPEFAKRNVKLIALSIDSVEDHLAWSKDINAYNCEEPTEKLPFPI
                     IDDRNRELAILLGMLDPAEKDEKGMPVTARVVFVFGPDKKLKLSILYPATTGRNFDEI
                     LRVVISLQLTAEKRVATPVDWKDGDSVMVLPTIPEEEAKKLFPKGVFTKELPSGKKYL
                     RYTPQP"
     gene            join(7066. .7349,11228. .11374,12122. .12268,13602. .13882)
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-003"
     mRNA            join(7066. .7349,11228. .11374,12122. .12268,13602. .13882)
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-003"
                     /product="peroxiredoxin 6"
                     /note="match: ESTs: Em:AV751940.1 Em:BG025204.1"
     polyA_site      13882
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-003"
     polyA_signal    14650. .14655
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-001"
     polyA_site      14674
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-001"
     gene            complement(join(26331. .26964,29133. .29193,31114. .31198,
                     32723. .32908,35435. .35566,43404. .43571,47168. .47266,
                     49836. .49961,50646. .50794,52530. .52581,55772. .55921,
                     59476. .59664,60351. .60549,61697. .61816,62809. .62906,
                     73556. .73651,74256. .74419,80600. .80681,83207. .83375,
                     99049. .99192,102528. .102627,107714. .107875,109373. .109489,
                     113532. .113697,123771. .123899,125984. .126084,
                     127517. .127722,128618. .128961))
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-001"
     mRNA            complement(join(26331. .26964,29133. .29193,31114. .31198,
                     32723. .32908,35435. .35566,43404. .43571,47168. .47266,
                     49836. .49961,50646. .50794,52530. .52581,55772. .55921,
                     59476. .59664,60351. .60549,61697. .61816,62809. .62906,
                     73556. .73651,74256. .74419,80600. .80681,83207. .83375,
                     99049. .99192,102528. .102627,107714. .107875,109373. .109489,
                     113532. .113697,123771. .123899,125984. .126084,
                     127517. .127722,128618. .128961))
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-001"
                     /product="novel protein"
                     /note="match: cDNAs: Em:AK128104.1 Em:BC042592.1"
     polyA_site      complement(26331)
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-001"
     gene            complement(join(26337. .26964,29133. .29203,31114. .31175,
                     52938. .53293,55772. .55921,59476. .59664,60351. .60385,
                     61697. .61816,62825. .62906,73556. .73651,74256. .74419,
                     80600. .80681,83207. .83375,99049. .99192,102528. .102627,
                     107714. .107785))
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-002"
     mRNA            complement(join(26337. .26964,29133. .29203,31114. .31175,
```

```
                     52938. .53293,55772. .55921,59476. .59664,60351. .60385,
                     61697. .61816,62825. .62906,73556. .73651,74256. .74419,
                     80600. .80681,83207. .83375,99049. .99192,102528. .102627,
                     107714. .107785))
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-002"
                     /product="novel protein"
                     /note="match: cDNAs: Em:AB072793.1"
     gene            complement(join(26337. .26964,29133. .29193,31114. .31198,
                     32723. .32908,35435. .35566,43404. .43571,47168. .47266,
                     49836. .49961,50646. .50794,52530. .52581,55772. .55921,
                     59476. .59664,60351. .60549,61697. .61816,62809. .62906,
                     73556. .73651,74256. .74419,83207. .83375,97369. .97595,
                     99049. .99192,102528. .102618))
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-003"
     mRNA            complement(join(26337. .26964,29133. .29193,31114. .31198,
                     32723. .32908,35435. .35566,43404. .43571,47168. .47266,
                     49836. .49961,50646. .50794,52530. .52581,55772. .55921,
                     59476. .59664,60351. .60549,61697. .61816,62809. .62906,
                     73556. .73651,74256. .74419,83207. .83375,97369. .97595,
                     99049. .99192,102528. .102618))
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-003"
                     /product="novel protein"
                     /note="match: cDNAs: Em:BX648119.1"
     polyA_site      complement(26337)
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-003"
     polyA_signal    complement(26352. .26357)
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-003"
     CDS             complement(join(26961. .26964,29133. .29193,31114. .31198,
                     32723. .32908,35435. .35566,43404. .43571,47168. .47266,
                     49836. .49961,50646. .50794,52530. .52581,55772. .55921,
                     59476. .59664,60351. .60549,61697. .61816,62809. .62906,
                     73556. .73651,74256. .74419,80600. .80681,83207. .83375,
                     99049. .99192,102528. .102627,107714. .107875,109373. .109489,
                     113532. .113697,123771. .123899,125984. .126084,
                     127517. .127643))
                     /gene="RP3-436N22.2"
                     /locus_tag="RP3-436N22.2-001"
                     /standard_name="OTTHUMP00000032688"
                     /note="match: proteins: Tr:Q86UF3 Tr:Q95LJ5"
                     /codon_start=1
                     /product="novel protein"
                     /protein_id="CAI20938.1"
                     /db_xref="GI:56204404"
                     /db_xref="GOA:Q5TAH2"
                     /db_xref="InterPro:IPR000595"
                     /db_xref="InterPro:IPR005820"
                     /db_xref="UniProt/TrEMBL:Q5TAH2"
                     /translation="MSSYFWAQNESNRPDLLCGQPADYLVEEKHFTTLVCFIVVLGGL
                     LKMCLKNCEVIVLTILSLSGFVIGHMAYNSVEVHQIVYPLLRTSSFSLYSYFSPLIIF
                     MVALDVEFYTLKKMFWQVLLTGLISFSTASIIIGYVVIKFNKDSWDLQSCLLFSITLG
                     IIDPLRSVNSLKTIGISKIYIDLIRGESLIICSIASIFFGNFRGNRIHFSIFRDLHVG
                     IELSYDILGSIIFGYWCAKIIQCILADVFSNMLTNIILCFSMVYMTFYIVEFLGMSGT
```

```
  Query Match             25.3%;  Score 505.2;  DB 8;  Length 159771;
  Best Local Similarity   74.3%;  Pred. No. 1.2e-69;
  Matches  780;  Conservative    0;  Mismatches  238;  Indels   32;  Gaps   10;

Qy        965 AGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACA 1024
              || | |  | |   |||||||| |||||||||| ||  ||||   ||||| |||||    |
Db      79697 AGAGATTTACATATCAGGCATAACTTGGAGATATTGCAGGTTCAATTCCAGACCACGTGA 79638

Qy       1025 ATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGC 1084
                ||   ||  |||||| ||    |||||||||||||  ||||| |||| |  | |||
Db      79637 CAAAGCAAATAATGCAATAA----ATATCACAATAAAGCAAGTCATACAAATTATATGGT 79582

Qy       1085 TTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA---A 1141
              |||||||| |||||||||||| || ||| |||| |||||||||  |||         |
Db      79581 TTCCCAGTACATATAAAAGTTGTGTTTACACTATACTGTAGTCAATTATAATCAAAAGCA 79522
```

```
Qy       1142 TAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACT-AAAAAAT 1200
              |||||        ||  |   |||||||||||   |||||||||| || || |||||||
Db      79521 CAGTGTCTAAGAAAACTATTTACATACCTTAATGAAAAAATGCTTTCTTGCTAAAAAAAT 79462

Qy       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260
              |||||||||||| ||||| |||||  | |  |||||||||||||||||  | |||   |
Db      79461 GCTAACAATCATCTGAGCCTTCAGCAATTCATAATCTTTTTGCTGGTGCAGGGTTGATAG 79402

Qy       1261 TTATTGATGACTGA---TCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTC 1317
              |  |||     | |     ||| |   | ||||||||  || ||  |      |  | ||
Db      79401 CTGCTGACTGATCAGATCAGGGTGGTGGTTGCTGAAGATTGGGCAGCTGTGGCAATTTCT 79342

Qy       1318 TTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCT 1377
              |||| |||||||||||| |||  | ||||  |||||||||||||||||||||||||||||
Db      79341 TTAAGACAACAGTGAAGTTTGGCACATCAACTGACTCTTCCTTTCATGAAAGATTTCTCT 79282

Qy       1378 CTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG- 1436
              |||  ||   |||| ||||||||||||||||   | ||||||  |||||||  ||||||||
Db      79281 GTAGCATGCAGTGCTGTTTGATAGCATTTTACATAGAGTAGAGCTTCTTTCAAAATTGGA 79222

Qy       1437 ATCAATCCTCTCAAACCC---TGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGA 1493
              || |||||||||||| ||  |||       || |  | |||||||  ||| ||||||||  |
Db      79221 ATAAATCCTCTCAAAACCTGTTGCCACTACTTTAGCAACTAAGTTTATAAAATATTCTAA 79162

Qy       1494 ATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTC 1553
              |||| ||| |||| |||||||||| |||||||  |||||||||||||||||| |||| ||
Db      79161 ATCCTTTGCTGTCCTTTCAACAATGTTCACAGCATCTTCACCAGGAGTAGATGCCATATC 79102

Qy       1554 ATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTT 1613
              |       | | |    ||| ||||||  ||||||| || ||||||||  | |||||||
Db      79101 AAGAAAGCACTTCG----TTTACTCATCTGTAAGAAGCAACTCCTCATCCATGCAAGTTT 79046

Qy       1614 TATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAG 1673
              |||| |||||||||||||||| ||||||  ||||||||      ||  |||| ||||||
Db      79045 TATCCTGAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCTACTTCTAATTCTAG 78991

Qy       1674 TTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTC 1733
              ||| |||||| |||| |||||||||  ||| |||||||| |||||||||||||| |||
Db      78990 TTCCCTTGCTATTTCCACCACATCTGCAGTTACTTCCTCCACTGAAGTCTTGAACCCCTC 78931

Qy       1734 CAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTT 1793
              |||||||| |||| ||||  |||| |||||||||  | |||| ||||| || |||||||
Db      78930 AAAGTCATCTATGAAGGTTAAAATCAACTTCTTCCTGACTCCTATTAATGTTGATATTTT 78871

Qy       1794 GACCTC---CCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
              ||||||   | |||||| |||||| |||||||||||||| ||||||||||  || ||||
Db      78870 GACCTCCTGCTATGAATTATGAATGCTCTTAATGGCACCTAGAATGGTGAATTCTATCCA 78811

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910
              |||||||  ||| | ||   | | ||||||     |||||||  |||| || ||| ||
Db      78810 TGAGGTTTTTTATTGATTTTCCCTACATCCAT-----CAGAGGAATCACTATCTATGACA 78756

Qy       1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970
              | ||||||||||   ||||| ||| || ||||||||||||||||||  |||| ||||||
Db      78755 GCTATAGCCTTACAAAATGTGTTTATTAAATAATAAGGCTTGAAAGTCAAAATGACTCCT 78696

Qy       1971 TGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||| | |   ||||| ||| |||||||||
Db      78695 TGATACTTGGGCTGCAGAATGGATGTTGTG 78666
```

```
RESULT 38
AC021635
LOCUS       AC021635              162163 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-418M15 map 3, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC021635
VERSION     AC021635.3  GI:7230225
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 162163)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-418M15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162163)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 162163)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6899683.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5968
                Center clone name: 418_M_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 158288 bases at least Q40
                Consensus quality: 160068 bases at least Q30
                Consensus quality: 160845 bases at least Q20
                Insert size: 163000; agarose-fp
                Insert size: 161363; sum-of-contigs
                Quality coverage: 5.9 in Q20 bases; agarose-fp
                Quality coverage: 6.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3780: contig of 3780 bp in length
            *     3781     3880: gap of 100 bp
            *     3881     7396: contig of 3516 bp in length
            *     7397     7496: gap of 100 bp
            *     7497    14714: contig of 7218 bp in length
            *    14715    14814: gap of 100 bp
            *    14815    31890: contig of 17076 bp in length
            *    31891    31990: gap of 100 bp
            *    31991    54940: contig of 22950 bp in length
            *    54941    55040: gap of 100 bp
            *    55041    79944: contig of 24904 bp in length
            *    79945    80044: gap of 100 bp
            *    80045   104673: contig of 24629 bp in length
            *   104674   104773: gap of 100 bp
            *   104774   131627: contig of 26854 bp in length
            *   131628   131727: gap of 100 bp
            *   131728   162163: contig of 30436 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162163
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-418M15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .3780
                     /note="assembly_fragment"
     gap             3781. .3880
                     /estimated_length=100
     misc_feature    3881. .7396
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     gap             7397. .7496
                     /estimated_length=100
     misc_feature    7497. .14714
                     /note="assembly_fragment"
     gap             14715. .14814
                     /estimated_length=100
     misc_feature    14815. .31890
                     /note="assembly_fragment"
     gap             31891. .31990
                     /estimated_length=100
     misc_feature    31991. .54940
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     gap             54941. .55040
                     /estimated_length=100
     misc_feature    55041. .79944
                     /note="assembly_fragment"
     gap             79945. .80044
                     /estimated_length=100
     misc_feature    80045. .104673
                     /note="assembly_fragment"
     gap             104674. .104773
                     /estimated_length=100
     misc_feature    104774. .131627
                     /note="assembly_fragment"
     gap             131628. .131727
                     /estimated_length=100
```

```
     misc_feature    131728. .162163
                     /note="assembly_fragment"
ORIGIN

  Query Match             25.3%;  Score 505.2;  DB 14;  Length 162163;
  Best Local Similarity   74.3%;  Pred. No. 1.2e-69;
  Matches  780;  Conservative    0;  Mismatches  238;  Indels   32;  Gaps   10;

Qy        965 AGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACA 1024
              || | |  | |   |||||||| |||||||||| ||  ||||   ||||| |||||   |
Db     101233 AGAGATTTACATATCAGGCATAACTTGGAGATATTGCAGGTTCAATTCCAGACCACGTGA 101292

Qy       1025 ATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGC 1084
                ||   ||  |||||| ||    |||||||||||||  ||||| |||| |   | |||
Db     101293 CAAAGCAAATAATGCAATAA----ATATCACAATAAAGCAAGTCATACAAATTATATGGT 101348

Qy       1085 TTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA---A 1141
              |||||||| |||||||||||| || ||| |||| |||||||||  |||        |   |
Db     101349 TTCCCAGTACATATAAAAGTTGTGTTTACACTATACTGTAGTCAATTATAATCAAAAGCA 101408

Qy       1142 TAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACT-AAAAAAT 1200
              |||||       ||  |   ||||||||||||   |||||||||| || || |||||||
Db     101409 CAGTGTCTAAGAAAACTATTTACATACCTTAATGAAAAAATGCTTTCTTGCTAAAAAAAT 101468

Qy       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260
              |||||||||||| ||||| |||||  | |  ||||||||||||||||| | |||   |
Db     101469 GCTAACAATCATCTGAGCCTTCAGCAATTCATAATCTTTTTGCTGGTGCAGGGTTGATAG 101528

Qy       1261 TTATTGATGACTGA---TCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTC 1317
              |  |||     | |     ||| |   | ||||||| || ||  |     |  | ||
Db     101529 CTGCTGACTGATCAGATCAGGGTGGTGGTTGCTGAAGATTGGGCAGCTGTGGCAATTTCT 101588

Qy       1318 TTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCT 1377
              |||| |||||||||||| |||  | ||||  |||||||||||||||||||||||||||||
Db     101589 TTAAGACAACAGTGAAGTTTGGCACATCAACTGACTCTTCCTTTCATGAAAGATTTCTCT 101648

Qy       1378 CTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG- 1436
                |||  ||   |||| ||||||||||||||||    | |||||| |||||| ||||||||
Db     101649 GTAGCATGCAGTGCTGTTTGATAGCATTTTACATAGAGTAGAGCTTCTTTCAAAATTGGA 101708

Qy       1437 ATCAATCCTCTCAAACCC---TGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGA 1493
              || ||||||||||| ||    |||      || |  | |||||| ||| |||||||| |
Db     101709 ATAAATCCTCTCAAAACCTGTTGCCACTACTTTAGCAACTAAGTTTATAAAATATTCTAA 101768

Qy       1494 ATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTC 1553
              |||| ||| |||| |||||||||| |||||||  |||||||||||||||||| |||| ||
Db     101769 ATCCTTTGCTGTCCTTTCAACAATGTTCACAGCATCTTCACCAGGAGTAGATGCCATATC 101828

Qy       1554 ATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTT 1613
              |       |  | |    ||| |||||  |||||| || |||||||  | |||||||
Db     101829 AAGAAAGCACTTCG----TTTACTCATCTGTAAGAAGCAACTCCTCATCCATGCAAGTTT 101884

Qy       1614 TATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAG 1673
              |||| ||||||||||||||| ||||||  ||||||||     ||  |||| |||||| ||
Db     101885 TATCCTGAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCTACTTCTAATTCTAG 101939

Qy       1674 TTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCCTCC 1733
              ||| |||||| |||| ||||||||||  ||| |||||||| |||||||||||||| |||
Db     101940 TTCCCTTGCTATTTCCACCACATCTGCAGTTACTTCCTCCACTGAAGTCTTGAACCCCTC 101999

Qy       1734 CAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTT 1793
              |||||||| |||| ||||  |||| |||||||||  | |||| |||| || |||||||
Db     102000 AAAGTCATCTATGAAGGTTAAAATCAACTTCTTCCTGACTCCTATTAATGTTGATATTTT 102059

Qy       1794 GACCTC---CCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
              ||||||   | |||||| ||||||| |||||||||||||| |||||||||| || ||||
Db     102060 GACCTCCTGCTATGAATTATGAATGCTCTTAATGGCACCTAGAATGGTGAATTCTATCCA 102119

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910
              |||||||  ||| | ||   | | ||||||     |||||||  |||| || ||| ||
Db     102120 TGAGGTTTTTTATTGATTTTCCCTACATCCAT-----CAGAGGAATCACTATCTATGACA 102174

Qy       1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970
              | ||||||||||   ||||| ||| || ||||||||||||||||||  |||| ||||||
Db     102175 GCTATAGCCTTACAAAATGTGTTTATTAAATAATAAGGCTTGAAAGTCAAAATGACTCCT 102234

Qy       1971 TGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||| | |   ||||| ||| |||||||||
Db     102235 TGATACTTGGGCTGCAGAATGGATGTTGTG 102264


RESULT 39
AL929203/c
LOCUS       AL929203              169424 bp    DNA     linear   HTG 01-OCT-2002
DEFINITION  Homo sapiens chromosome 1 clone RP11-258J6.
ACCESSION   AL929203
VERSION     AL929203.3  GI:23476799
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 169424)
  AUTHORS   Martin,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 2, 2002 this sequence version replaced gi:23395907.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA258J6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 166627 bases at least Q40
            Consensus quality: 166849 bases at least Q30
            Consensus quality: 166945 bases at least Q20
            Insert size: 169424; sum-of-contigs
            Insert size: 172092; 4.2% error; agarose-fp
            Quality coverage: 9.31x in Q20 bases; sum-of-contigs Quality
            coverage: 9.16x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   169424: contig of 169424 bp in length.
FEATURES              Location/Qualifiers
     source           1. .169424
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="1"
                      /clone="RP11-258J6"
                      /clone_lib="RPCI-11.1"
     misc_feature     1. .169424
                      /note="assembly_fragment:01428"
ORIGIN

  Query Match             25.3%;  Score 505.2;  DB 14;  Length 169424;
  Best Local Similarity   74.3%;  Pred. No. 1.2e-69;
  Matches  780;  Conservative    0;  Mismatches  238;  Indels   32;  Gaps   10;

Qy        965 AGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACA 1024
              || | |  | |   |||||||| |||||||||| ||  ||||   ||||| |||||   |
```

```
Db     108703 AGAGATTTACATATCAGGCATAACTTGGAGATATTGCAGGTTCAATTCCAGACCACGTGA 108644

Qy       1025 ATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGC 1084
              ||   ||  |||||| ||     ||||||||||||||  ||||| |||| |  | |||
Db     108643 CAAAGCAAATAATGCAATAA----ATATCACAATAAAGCAAGTCATACAAATTATATGGT 108588

Qy       1085 TTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA---A 1141
              |||||||| |||||||||||| || ||| |||| |||||||||  |||       |   |
Db     108587 TTCCCAGTACATATAAAAGTTGTGTTTACACTATACTGTAGTCAATTATAATCAAAAGCA 108528

Qy       1142 TAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACT-AAAAAAT 1200
              |||||       ||  |   ||||||||||||   |||||||||| || || |||||||
Db     108527 CAGTGTCTAAGAAAACTATTTACATACCTTAATGAAAAAATGCTTTCTTGCTAAAAAAAT 108468

Qy       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260
              |||||||||||| ||||| |||||  | |  ||||||||||||||||| | |||   |
Db     108467 GCTAACAATCATCTGAGCCTTCAGCAATTCATAATCTTTTTGCTGGTGCAGGGTTGATAG 108408

Qy       1261 TTATTGATGACTGA---TCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTC 1317
              |  |||     | |     ||| |   | ||||||| || ||  |     |  | ||
Db     108407 CTGCTGACTGATCAGATCAGGGTGGTGGTTGCTGAAGATTGGGCAGCTGTGGCAATTTCT 108348

Qy       1318 TTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCT 1377
              |||| |||||||||||| |||  | ||||  |||||||||||||||||||||||||||||
Db     108347 TTAAGACAACAGTGAAGTTTGGCACATCAACTGACTCTTCCTTTCATGAAAGATTTCTCT 108288

Qy       1378 CTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG- 1436
              |||  ||   |||| |||||||||||||||   | |||||| ||||||| ||||||||
Db     108287 GTAGCATGCAGTGCTGTTTGATAGCATTTTACATAGAGTAGAGCTTCTTTCAAAATTGGA 108228

Qy       1437 ATCAATCCTCTCAAACCC---TGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGA 1493
              || |||||||||||| ||   |||       || |  | ||||||| ||| |||||||| |
Db     108227 ATAAATCCTCTCAAAACCTGTTGCCACTACTTTAGCAACTAAGTTTATAAAATATTCTAA 108168

Qy       1494 ATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTC 1553
              |||| ||| |||| |||||||||| |||||||  ||||||||||||||||||| |||| ||
Db     108167 ATCCTTTGCTGTCCTTTCAACAATGTTCACAGCATCTTCACCAGGAGTAGATGCCATATC 108108

Qy       1554 ATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTT 1613
              |     |  | |    ||| |||||| ||||||| || |||||||  | |||||||
Db     108107 AAGAAAGCACTTCG----TTTACTCATCTGTAAGAAGCAACTCCTCATCCATGCAAGTTT 108052

Qy       1614 TATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAG 1673
              |||| |||||||||||||| |||||| |||||||     ||  |||| |||||| ||
Db     108051 TATCCTGAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCTACTTCTAATTCTAG 107997

Qy       1674 TTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTC 1733
              ||| |||||| |||| |||||||||   ||| ||||||||| ||||||||||||| |||
Db     107996 TTCCCTTGCTATTTCCACCACATCTGCAGTTACTTCCTCCACTGAAGTCTTGAACCCCTC 107937

Qy       1734 CAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTT 1793
              ||||||||  |||| ||||  |||| |||||||||  | |||| |||||| || |||||||
Db     107936 AAAGTCATCTATGAAGGTTAAAATCAACTTCTTCCTGACTCCTATTAATGTTGATATTTT 107877

Qy       1794 GACCTC---CCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
              ||||||    | |||||| |||||||  ||||||||||||||| ||||||||||| || ||||
Db     107876 GACCTCCTGCTATGAATTATGAATGCTCTTAATGGCACCTAGAATGGTGAATTCTATCCA 107817

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910
                |||||||  ||| | ||   | | ||||||     |||||||  |||| || ||| ||
Db     107816 TGAGGTTTTTTATTGATTTTCCCTACATCCAT-----CAGAGGAATCACTATCTATGACA 107762

Qy       1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970
              | ||||||||||   ||||| ||| || ||||||||||||||||||  |||| ||||||
Db     107761 GCTATAGCCTTACAAAATGTGTTTATTAAATAATAAGGCTTGAAAGTCAAAATGACTCCT 107702

Qy       1971 TGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||| | |   ||||| ||| |||||||||
Db     107701 TGATACTTGGGCTGCAGAATGGATGTTGTG 107672

RESULT 40
CQ870480
```

```
LOCUS       CQ870480              188759 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 901 from Patent WO2004074320.
ACCESSION   CQ870480
VERSION     CQ870480.1  GI:52000009
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Morris,D.W., Morris,D.W. and Malandro,M.S.
  TITLE     Novel therapeutic targets in cancer
  JOURNAL   Patent: WO 2004074320-A 901 02-SEP-2004;
            Sagres Discovery, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .188759
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="seq 638, original 788.759 bases, replaced by
                     seq 638, from 0.000.001 to 0.349.980
                     seq 900, from 0.300.001 to 0.649.980
                     seq 901, from 0.600.001 to 0.788.759"
ORIGIN

  Query Match           25.3%;  Score 505.2;  DB 6;  Length 188759;
  Best Local Similarity   75.5%;  Pred. No. 1.1e-69;
  Matches  743;  Conservative    0;  Mismatches  208;  Indels   33;  Gaps    8;

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
              || |  |  ||||| ||||  |||||||    |||| ||  | | |||||| || ||||||
Db     104541 ATCTCAATAAAGTGAATATTGCAATAAATAAAGTCGCAGGAATTTTTTGGTTTTCCAGTG 104600

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
              |||||||||||| || ||||| || |||||||||| |||||| | |||||||  |||||||
Db     104601 CATATAAAAGTTATGTTTATATTATACTGTAGTCTATTAAGTATTCAATAGCATTATGTC 104660

Qy       1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
              |||||||          ||| | ||| |||      |  | || |||| | |||||   ||
Db     104661 TAAAAAATATACATACTTTAATTTTAAAGTACCTTATGCTCAAAAGTACTAACCCCCACC 104720

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT---------TTTCTTAT 1264
              ||||| |||||||||||||||||||||||||||| ||| |          ||  |
Db     104721 TGAGCCTTCAGTGAGTTGTAATCTTTTTGCTGGTGGAGGGTTTGCCTCAAAGTTGATGTC 104780

Qy       1265 TGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-- 1322
              |  ||||||||| | |       |||| ||| |  | ||||||||||||||||||||
Db     104781 TCCTGACTGATCAGAGTGGTATTGCTAAAGGTAGGAGTGGCTGTGGCAGTTTCTTAAAAT 104840

Qy       1323 ---ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
                  ||||||||||  |||| | |||| |||||||||| |||||||||||| |||||| |
Db     104841 GAGACAACAGTGAAACTTGCCACATCAATTGACTCTTCTTTTCATGAAAGAATTCTCTGT 104900

Qy       1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-T 1438
              ||  ||||||| | |||| || |||||||  |||||||||||||| ||||  ||||| |
Db     104901 AGCATGTGATGTTGTTTGGTACCATTTTACTCACAGTAGAACTTCTGTGAACGTTGGAGT 104960

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
              ||  | |||||||||| |||       || |  | |||  |  ||  |||||||| || |
Db     104961 CACACTTCTCAAACCTTGCCACTTCTTCATCAACTATATGTATGGAATATTCTAAACCTT 105020

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
              |||||||||||||||||| |||| ||| | |||||||||||||| |||| ||||||
Db     105021 TTGTTGTCATTTCAACAATGTTCATAGTATTTTCACCAGGAGTATATTCTATCTCAAGAA 105080

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
              |  |     |||||||||||  ||||||  || |  ||| |  || |||||  ||||
Db     105081 GCCACTT----TCTTTGCTCATCTGTAAGAAACAAATGTTCACCCATTTAAGTTCGATCA 105136

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
              ||| |||||||||| ||||||   ||||||||      || |  || |||||| |||||||
Db     105137 TGACATTGCAGCAATTCAGTCTCATCTTCAGG-----CTCCCTTTCTAATTCTAGTTCTC 105191
```

```
Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
             ||||| ||||||||||||| ||  ||| ||||||||| ||||||||||||||||||| ||||
Db     105192 TTGCTATTTCTACCACATATGCAGTTACTTCCTCCACTGAAGTCTTGAACCTCTCAAAGT 105251

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
             ||||||| |||||||||||| | ||||| | || ||||||| || || ||||||||||||
Db     105252 CATCCATCAGGGTTGGAATCAATTTCTTGCCAACTCCTGTTTATGTTGATATTTTGACCT 105311

Qy       1799 CC---CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855
              ||   ||||||| |  | | |||| |||||||||| |||||||||||||||| || ||||
Db     105312 CCTCTCATGAATTACAAGTATTCTGAATGGCACCTAGAATGGTGAATCCTTTTCAGAAGG 105371

Qy       1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915
              |||||||||||||| | ||||||||||      |||| ||  |||| || ||| ||| ||
Db     105372 -TTTCAATTTACTTTGCCCAGATCCAT-----CAGAAGAATCACTATCTATGGCAGCTAC 105425

Qy       1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
              ||||||||| | |||||||||| ||| ||||  |||||||||  ||||||  ||||||||
Db     105426 AGCCTTATGAAGTGTATTTCTTAAATGATAAAACTTGAAAGTCAAAATTATACCTTGATC 105485

Qy       1976 CATTTTCTGCAAAATAGATGTTGT 1999
              |||   ||||| ||| ||||||||
Db     105486 CATGGGCTGCAGAATGGATGTTGT 105509


RESULT 41
CNS07EG5
LOCUS       CNS07EG5              175210 bp    DNA     linear   PRI 25-JUN-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-818K5 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL590328
VERSION     AL590328.2  GI:14571668
KEYWORDS    HTG; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 175210)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175210)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Jun 26, 2001 this sequence version replaced gi:13509295.
            -------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            ---------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            -------------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.46x in Q20 bases; sum-of-contigs
            ---------------
            ----------------------
            Overall quality chart :
            Range    :   bases
            0        :       6
            1  -  9  :      52
            10 - 19  :     268
            20 - 29  :     576
            30 - 39  :    1644
            40 - 49  :    7554
            50 - 59  :    9152
            60 - 69  :   10935
            70 - 79  :   21682
            80 - 89  :   43918
            90 - 99  :   79423
            ----------------------
            Percentage of bases with a quality value >= 40 : 98 %.
FEATURES             Location/Qualifiers
     source          1. .175210
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-818K5"
                     /clone_lib="RPCI-11"
ORIGIN

  Query Match             25.2%;  Score 504;  DB 8;  Length 175210;
  Best Local Similarity   75.0%;  Pred. No. 1.8e-69;
  Matches  784;  Conservative    0;  Mismatches  205;  Indels   57;  Gaps   10;

Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032
              | |||| |||||||| | |||||||||| |||| |||||| ||  ||||| ||| |
Db     164169 ACAGTATAGGCATACTTCGGAGATACTGTAGGTTCAGTTCCAGACTGCCACACTAAAA-- 164226

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092
                       || | ||  ||||||||   ||||||  ||| |||||| ||||||||
Db     164227 ----------TGAACATTGCAATAAAGCAGGTCACATGAGTTTTTTGGTTTCCCAGT 164273

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152
              ||||| ||||||| |  ||||     |||||||||  |||||||||||||||  ||||||
Db     164274 GCATACAAAAGTTATCTTTAT-----ACTGTAGTCCATTAAGTGTGCAATAGCATTATGT 164328

Qy       1153 CTAAAAAAACACAT-----ACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACA 1207
              |||||||||  ||       ||| | |||||||| |||||||  |||||||||| |||
Db     164329 CTAAAAAAAGTAATTTACATACTTTAATTTAAAATACTTTATTTCTAAAAAATGTTAATG 164388

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              ||||| ||||  |||||  |||||||||| |||||||| |||||||| ||  |||  ||
Db     164389 ATCATCTGAGTCTTCAGCAAGTTGTAATTTTTTTGCTGTTGGAAGGTTTTGCCTTGATGT 164448

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              |||  | |            ||| |    | |  ||||| ||   |||||||||||| |||
Db     164449 TGATGGTTGCTGAATGATCAGGGTGGTGGTTAGTGAAGGTTGAAGTGGCTGTGGCAATTT 164508

Qy       1317 CTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              ||||||     |||||||||||| | || | |||  |||||||||||||||  ||| |
Db     164509 CTTAAAATAAGACAACAGTGAAGTTGGCCACATCTATTGACTCTTCCTTTCACAAAAAAG 164568

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
              ||| || |||| |||||||| |||||||||||||||| |||||||||||||||| |||
Db     164569 TTCCCTGTAGTATGTGATGCTGTTTGATAGCATTTTATCCACAGTAGAACTTCTTTCAAA 164628

Qy       1432 ATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              |||||| || |||||||||||||||||  |   ||  | | ||||||| || ||||||||
Db     164629 ATTGGATTCCATCCTCTCAAACCCTGCTGTTGCTTTTC-AGCTAAGTTTATGTAATATTC 164687

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
              |||||||  ||   |||||||||||  |||||||  |||||||||||||||||||||||
Db     164688 TGAATCCTCTGAATTCATTTCAACAACGTTCACAGCATCTTCACCAGGAGTAGATTCCAT 164747

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              |||    |   |  |  ||||||||||   | ||||  || |||| |||||||||
Db     164748 CTC----CAGAAACTACTTTCTTTGCTCATGTGTGAGAATCAACTCCTTATCTGTTCATA 164803

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              || |||||||||| ||||| || || |    ||||||||       | ||||| ||||||
Db     164804 TTGTATCATGAGATGGCAGCCATTCAATTGCATCTTCAGG-----ATCCACTTCTAATTC 164858

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
              |||||||||||||||| ||||||||||   || |||||||    ||| | ||||| |
Db     164859 TAGTTCTCTTGCTGTTTCCACCACATCTGCAATTACTTCCTCTGCCGAAATTTTGAATCC 164918
```

```
Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
             ||| |||||| |||||||| || ||||  |||||||||| ||||||| |||| |   | |
Db     164919 CTCAAAGTCAGCCATGAGGATTAGAATAAACTTCTTCCAGATTCCTGCTAATGTCCTTTT 164978

Qy       1791 TTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
             |||||||| |||| ||||||||| |||||||||||| ||  |||  |||||||||||| |
Db     164979 TTTGACCTTTCATGCATCATGAATTTTCTTAATGGCATCTAAAATCCTGAATCCTTTCTA 165038

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910
             ||| ||||||||||   || | |||| |||||     |||||||  |||| || ||| ||
Db     165039 AAAAGTTTTCAATTGCTTTTGCCCAGGTCCAT-----CAGAGGAATCACTATCTATGGCA 165093

Qy       1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970
             || |||||||||   ||| || || || | ||||||  ||||||||||||||||||||
Db     165094 GTGATAGCCTTACAAAATTTAGTTTTTAAGTAATAACACTTGAAAGTTGAAATTACTCCC 165153

Qy       1971 TGATCCATTTTCTGCAAAATAGATGT 1996
             ||||||||    |||| ||| |||||
Db     165154 TGATCCATGGGATGCAGAATGGATGT 165179


RESULT 42
CNS01DSQ
LOCUS       CNS01DSQ              200853 bp    DNA     linear   PRI 26-SEP-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-747H7 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL121839
VERSION     AL121839.3  GI:15796542
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 200853)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200853)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Sep 27, 2001 this sequence version replaced gi:9955595.
            --------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            ---------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-818K5 (AC=AL590328)
            Downstream BAC (overlapping the SP6 end) : R-248B24 ----------
            Summary Statistics
            Assembly program: Phrap; version 2.0
            ---------------
            ----------------------
            Overall quality chart :
            Range    :   bases
            0        :
            1  -  9  :       1
            10 - 19  :      10
            20 - 29  :      37
            30 - 39  :     228
            40 - 49  :    3790
            50 - 59  :   12266
            60 - 69  :   14881
            70 - 79  :   23014
            80 - 89  :   49600
            90 - 99  :   97026
            ----------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES             Location/Qualifiers
     source          1. .200853
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-747H7"
                     /clone_lib="RPCI-11"
     STS             74706. .74805
                     /note="matching EMBL:R33632
                     RHdb:RH53920
                     dbSTS:STS32121
                     Identified using the e-PCR software (G. Schuler)"
     STS             82261. .82445
                     /note="matching EMBL:R25802
                     RHdb:RH44322
                     RHdb:RH11302
                     dbSTS:STS1025
                     Identified using the e-PCR software (G. Schuler)"
     STS             83496. .83749
                     /note="matching EMBL:H09564
                     RHdb:RH53633
                     dbSTS:STS42877
                     Identified using the e-PCR software (G. Schuler)"
     STS             119346. .119486
                     /note="matching EMBL:G09073
                     RHdb:RH53763
                     RHdb:RH28183
                     dbSTS:STS32298
                     Identified using the e-PCR software (G. Schuler)"
ORIGIN

  Query Match           25.2%;  Score 504;  DB 8;  Length 200853;
  Best Local Similarity   75.0%;  Pred. No. 1.7e-69;
  Matches  784;  Conservative    0;  Mismatches  205;  Indels   57;  Gaps   10;

Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032
              | |||| |||||||| | |||||||||||| ||||  |||||| ||  ||||| ||| |
Db      60151 ACAGTATAGGCATACTTCGGAGATACTGTAGGTTCAGTTCCAGACTGCCACACTAAAA-- 60208

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092
                    || | ||  |||||||||   ||||||  ||| |||||| ||||||||
Db      60209 -------------TGAACATTGCAATAAAGCAGGTCACATGAGTTTTTTGGTTTCCCAGT 60255

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152
              ||||| ||||||| | ||||      |||||||||  ||||||||||||||  ||||||
Db      60256 GCATACAAAAGTTATCTTTAT-----ACTGTAGTCCATTAAGTGTGCAATAGCATTATGT 60310

Qy       1153 CTAAAAAAACACAT-----ACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACA 1207
              |||||||||   ||       ||| | |||||||| |||||||  |||||||||| |||
Db      60311 CTAAAAAAAGTAATTTACATACTTTAATTTAAAATACTTTATTTCTAAAAAATGTTAATG 60370

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              ||||| ||||  |||||  |||||||||| |||||||| |||||||| ||  |||  ||
Db      60371 ATCATCTGAGTCTTCAGCAAGTTGTAATTTTTTTGCTGTTGGAAGGTTTTGCCTTGATGT 60430

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              |||  | |            ||| |   | |  ||||| ||   ||||||||||| |||
Db      60431 TGATGGTTGCTGAATGATCAGGGTGGTGGTTAGTGAAGGTTGAAGTGGCTGTGGCAATTT 60490

Qy       1317 CTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              ||||||     |||||||||||| | || | |||  ||||||||||||||   ||| |
Db      60491 CTTAAAATAAGACAACAGTGAAGTTGGCCACATCTATTGACTCTTCCTTTCACAAAAAAG 60550

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
              ||| || |||| ||||||||| |||||||||||||||| |||||||||||||||| |||
Db      60551 TTCCCTGTAGTATGTGATGCTGTTTGATAGCATTTTATCCACAGTAGAACTTCTTTCAAA 60610

Qy       1432 ATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              |||||| || |||||||||||||||||| |   ||  | | ||||||| || ||||||||
```

```
Db      60611 ATTGGATTCCATCCTCTCAAACCCTGCTGTTGCTTTTC-AGCTAAGTTTATGTAATATTC 60669

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
              ||||||| ||   |||||||||||  ||||||| ||||||||||||||||||||||||
Db      60670 TGAATCCTCTGAATTCATTTCAACAACGTTCACAGCATCTTCACCAGGAGTAGATTCCAT 60729

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              |||    |   |   |    ||||||||||    | ||||  || |||| |||||||||
Db      60730 CTC----CAGAAACTACTTTCTTTGCTCATGTGTGAGAATCAACTCCTTATCTGTTCATA 60785

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              || |||||||||| ||||| || || |    |||||||       | |||||  ||||||
Db      60786 TTGTATCATGAGATGGCAGCCATTCAATTGCATCTTCAGG-----ATCCACTTCTAATTC 60840

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
               ||||||||||||||||| ||||||||||   || |||||||    ||| | ||||| |
Db      60841 TAGTTCTCTTGCTGTTTCCACCACATCTGCAATTACTTCCTCTGCCGAAATTTTGAATCC 60900

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              ||| |||||| |||||||| || ||||  |||||||||| ||||||| |||| |    | |
Db      60901 CTCAAAGTCAGCCATGAGGATTAGAATAAACTTCTTCCAGATTCCTGCTAATGTCCTTTT 60960

Qy       1791 TTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
              ||||||||  |||| ||||||||| |||||||||||| ||  |||  |||||||||| |
Db      60961 TTTGACCTTTCATGCATCATGAATTTTCTTAATGGCATCTAAAATCCTGAATCCTTTCTA 61020

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910
              ||| ||||||||||   || | |||| |||||     |||||||  |||| || ||| ||
Db      61021 AAAAGTTTTCAATTGCTTTTGCCCAGGTCCAT-----CAGAGGAATCACTATCTATGGCA 61075

Qy       1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970
              || |||||||||   ||| || || || | ||||||  ||||||||||||||||||||||
Db      61076 GTGATAGCCTTACAAAATTTAGTTTTTAAGTAATAACACTTGAAAGTTGAAATTACTCCC 61135

Qy       1971 TGATCCATTTTCTGCAAAATAGATGT 1996
              ||||||||    |||| ||| |||||
Db      61136 TGATCCATGGGATGCAGAATGGATGT 61161
```

```
RESULT 43
AC004764
LOCUS       AC004764               68048 bp    DNA     linear   PRI 29-MAY-1998
DEFINITION  Homo sapiens chromosome 5, P1 clone 255g5 (LBNL H61), complete
            sequence.
ACCESSION   AC004764 AC001063 AC002245 AC002246 L81711 L81712 L81884 L81885
            L81886 L81887
VERSION     AC004764.1  GI:3168619
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 68048)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68048)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 68048)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .68048
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q"
                     /clone="255g5"
                     /note="LBNL H61"
     repeat_region   complement(187. .491)
                     /rpt_family="Alu"
     repeat_region   517. .552
                     /note="(GT)18"
                     /rpt_type=tandem
                     /rpt_unit="gt"
     repeat_region   complement(548. .1010)
                     /rpt_family="Alu"
     repeat_region   complement(2267. .2592)
                     /rpt_family="Alu"
     misc_feature    complement(3203. .3434)
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(5139. .5432)
                     /rpt_family="Alu"
     repeat_region   complement(5422. .5925)
                     /rpt_family="Alu"
     misc_feature    complement(5920. .6527)
                     /note="95% identity yx66g09.s1"
                     /db_xref="dbEST:N23052"
     repeat_region   6791. .7162
                     /rpt_family="Alu"
     repeat_region   complement(7294. .7362)
                     /rpt_family="MER42"
     repeat_region   8033. .8105
                     /rpt_family="MIR2"
     misc_feature    8384. .8537
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   complement(8593. .8928)
                     /rpt_family="Alu"
     repeat_region   complement(9004. .9331)
                     /rpt_family="Alu"
     repeat_region   9014. .9034
                     /note="(T)21"
                     /rpt_type=tandem
                     /rpt_unit="t"
     repeat_region   complement(9390. .9693)
                     /rpt_family="Alu"
     repeat_region   9710. .9884
                     /rpt_family="Alu"
     prim_transcript complement(join(9901. .9985,12737. .12991))
                     /standard_name="yyt67a05"
                     /note="95% & 99% identity EST dbEST:R84804 & R85570"
     misc_feature    10242. .10367
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   10545. .10779
                     /rpt_family="MER46"
     repeat_region   complement(10550. .10757)
                     /rpt_family="MER46"
     repeat_region   11943. .12174
                     /rpt_family="MER42"
     STS             complement(12737. .12991)
                     /standard_name="SHGC-30875"
                     /note="99% identity"
                     /db_xref="dbSTS:G29311"
     repeat_region   13888. .14069
                     /rpt_family="Alu"
     repeat_region   14056. .14502
                     /rpt_family="L1"
     repeat_region   complement(14536. .14660)
                     /rpt_family="MSTa"
     repeat_region   14687. .14787
                     /rpt_family="Alu"
     repeat_region   complement(14871. .16160)
```

```
                     /rpt_family="MSTAR"
     repeat_region   complement(16201. .16573)
                     /rpt_family="MSTa"
     repeat_region   16578. .16898
                     /rpt_family="L1MA5"
     repeat_region   16970. .17136
                     /rpt_family="L1"
     repeat_region   complement(17157. .17470)
                     /rpt_family="Alu"
     repeat_region   complement(17596. .17863)
                     /rpt_family="Alu"
     repeat_region   18183. .18338
                     /rpt_family="L1MB6"
     repeat_region   complement(18430. .18777)
                     /rpt_family="Alu"
     repeat_region   complement(18901. .19199)
                     /rpt_family="Alu"
     repeat_region   complement(19313. .19502)
                     /rpt_family="Alu"
     repeat_region   19680. .19986
                     /rpt_family="Alu"
     repeat_region   20240. .20302
                     /rpt_family="L1MB6"
     repeat_region   20299. .20538
                     /rpt_family="Alu"
     repeat_region   20955. .21167
                     /rpt_family="L1"
     repeat_region   21802. .22046
                     /rpt_family="Alu"
     repeat_region   22298. .23300
                     /rpt_family="L1"
     repeat_region   complement(23291. .23615)
                     /rpt_family="Alu"
     repeat_region   23588. .23979
                     /rpt_family="L1MA2"
     repeat_region   24086. .24388
                     /rpt_family="Alu"
     repeat_region   complement(24617. .25236)
                     /rpt_family="Alu"
     repeat_region   24929. .24949
                     /note="(T)21"
                     /rpt_type=tandem
                     /rpt_unit="t"
     repeat_region   25335. .26483
                     /rpt_family="L1"
     repeat_region   complement(26476. .28579)
                     /rpt_family="Tigger1"
     repeat_region   complement(28589. .29033)
                     /rpt_family="MER4"
     repeat_region   complement(29028. .29221)
                     /rpt_family="Tigger1"
     repeat_region   29445. .30152
                     /rpt_family="L1"
     repeat_region   30319. .30630
                     /rpt_family="Alu"
     repeat_region   complement(30604. .30887)
                     /rpt_family="RBL10C4"
     repeat_region   30898. .31218
                     /rpt_family="Alu"
     repeat_region   31213. .31918
                     /rpt_family="L1"
     repeat_region   32291. .32625
                     /rpt_family="Alu"
     repeat_region   complement(33507. .33877)
                     /rpt_family="L1"
     repeat_region   complement(35113. .35272)
                     /rpt_family="MIR"
     repeat_region   complement(36395. .42569)
                     /rpt_family="L1"
     repeat_region   complement(40993. .41144)
                     /rpt_family="MER25"
     repeat_region   45011. .45337
                     /rpt_family="Alu"
     repeat_region   complement(45637. .45703)
                     /rpt_family="MIR"
     misc_feature    46384. .46629
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    47310. .47464
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    47840. .47963
                     /note="GRAIL 2 excellent exon, frame 1"
     misc_feature    48883. .49040
                     /note="GRAIL 2 excellent exon, frame 0"
     prim_transcript complement(join(48972. .49041,50028. .50217))
                     /standard_name="ng41h11"
                     /note="99% identity dbEST:AA527520"
     misc_feature    50031. .50133
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(52346. .52761)
                     /rpt_family="THE1"
     repeat_region   54686. .54836
                     /rpt_family="Alu"
     repeat_region   56010. .56375
                     /rpt_family="Alu"
     prim_transcript join(57722. .57914,60189. .60588)

  Query Match            25.2%;  Score 503.4;  DB 8;  Length 68048;
  Best Local Similarity  77.7%;  Pred. No. 3.2e-69;
  Matches  736;  Conservative    0;  Mismatches  171;  Indels   40;  Gaps    9;

Qy       1051 ATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCT 1110
             | |||||||||||||||||||  | | |||| | ||||  || ||||||||||||  | |
Db      26515 ACCACAATAAAGTGAGTCACATGAATTTTTTTGTTTCCTGGTACATATAAAAGTTAAGGT 26574

Qy       1111 TATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC-TAAAAAAACACATACC 1169
              || | || |||||||||||||||||||||||||||  |||  || |||||||  | ||||
Db      26575 TACATTATACTGTAGTCTGTTAAGTGTGCAATAGCATTAGATCTTAAAAAAGTATATACT 26634

Qy       1170 TTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGT 1229
              ||||||| ||||| |||||||| |||||||||| ||||||||| ||||| |||||  ||
Db      26635 TTAATTTAAAAATACTTTATTGCTAAAAAATGCCAACAATCATCTGAGCCTTCAGCAAGA 26694

Qy       1230 TGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CGG 1278
                |||||||||||||||||| ||||||  |||  || |||  | |           ||
Db      26695 CATAATCTTTTTGCTGGTGGAGGGTCTTGCCTTGATGTTGATGGCTGCTAACTGATCAGG 26754

Qy       1279 GGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAA 1333
                |   | ||||||||  | | ||| |||||||| |||||||||     ||||||  |||
Db      26755 ATGATGGTTGCTGAAGGCTGGAGTGTCTGTGGCAATTTCTTAAAATAAGACAACAAGGAA 26814

Qy       1334 GATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTT 1393
              | |||| | |||| |||||||||||||||||||||||||||||| ||| |||   || |||||
Db      26815 GTTTGCCACATCAATTGACTCTTCCTTTCATGAAAGATTTTTCTATAGCAAGTAATGCTG 26874

Qy       1394 TTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAAC 1452
              |||| |   || |||  |||||||||||||||| ||| ||| |  ||||||||||||||
Db      26875 TTTGTTGAAATATTACCCACAGTAGAACTTCTTTCAAAGTTGAAGCCAATCCTCTCAAAC 26934

Qy       1453 CCTG-CTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTC 1511
              |||| | ||||||| |  | | ||||| |   |||||||| ||||| || ||||||||||
Db      26935 CCTGCCACTGCTTTTATCAACGAAGTTTACGGAATATTCTAAATCCTTTCTTGTCATTTC 26994

Qy       1512 AACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATC 1571
              ||||| ||||| ||  |||||  |||||  ||||||||||||||        || |    ||
Db      26995 AACAACTTTCATAGCATCTTCGCCAGGCGTAGATTCCATCTCAAGGAACCAGTT----TC 27050

Qy       1572 TTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCA 1631
              |||||||| |||| | ||| || |||||||||  ||||||||||||||||||||||||||||
Db      27051 TTTGCTCACCCATGAAAAGCAACTCCTCATCTCTTCAAGTTTTATCATGAGATTGCAGCA 27110

Qy       1632 ATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTAC 1691
                |  |||||  |||||| |     || || || |||||| |||||  |||||| |||| ||
Db      27111 GTTTAGTCACATCTTCAAG-----CTCCATTTCTAATTCTAGTTCACTTGCTATTTCCAC 27165

Qy       1692 CACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGT 1751
              |||||||   || ||||||||| |||||||||||||| ||| ||||||||||  ||||||
```

```
Db      27166 AACATCTGCAATTACTTCCTCCAGTGAAGTCTTGAACCCCTCAAAGTCATCCACGAGGGT 27225

Qy       1752 TGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATC 1808
              ||||||| || |||||||||||||||||| || || |||||||||   | |||  ||| ||
Db      27226 TGGAATCAACCTCTTCCAAATTCCTGTTCATGTTGATATTTTGATCTCTTCCTGTGAGTC 27285

Qy       1809 ATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACT 1868
              |||||||||||||||||| || | ||||||||| ||||||| |||||||||||||||||
Db      27286 CTGAATGTTCTTAATGGCATCTAGGATGGTGAATTCTTTCCAGAAGGTTTTCAATTTACT 27345

Qy       1869 TAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAAT 1928
              |   ||||| ||||||             | |  | ||| ||| ||||||||||| |||
Db      27346 TTACCCAGACCCATCC---------GAAGAATCACTATGTCAGCTATAGCCTTATGAAAT 27396

Qy       1929 GTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
              ||||||||| |||||||||  || |||||||||| ||||||||||
Db      27397 GTATTTCTTAAATAATAAGAGTTAAAAGTTGAAATGACTCCTTGATC 27443


RESULT 44
AC008670/c
LOCUS       AC008670              126034 bp    DNA     linear   PRI 28-JAN-2003
DEFINITION  Homo sapiens chromosome 5 clone CTB-36O1, complete sequence.
ACCESSION   AC008670
VERSION     AC008670.6  GI:27923608
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 126034)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 126034)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 126034)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 126034)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   5  (bases 1 to 126034)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jan 28, 2003 this sequence version replaced gi:15187197.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 2.1.
            NOTE:  This insert is not the entire sequence of the clone (entire
            sequence is 133kb). It is clipped at the overlap with AC006077. The
            number of bases overlapped is 36042.
FEATURES             Location/Qualifiers
     source          1. .126034
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-36O1"
ORIGIN


 Query Match           25.2%;  Score 503.4;  DB 8;  Length 126034;
 Best Local Similarity  77.7%;  Pred. No. 2.5e-69;
 Matches  736;  Conservative    0;  Mismatches  171;  Indels   40;  Gaps    9;

Qy       1051 ATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCT 1110
              | |||||||||||||||||||  | | |||| | ||||  || ||||||||||||  | |
Db      22210 ACCACAATAAAGTGAGTCACATGAATTTTTTTGTTTCCTGGTACATATAAAAGTTAAGGT 22151

Qy       1111 TATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC-TAAAAAAACACATACC 1169
              || | || ||||||||||||||||||||||||||||||  |||  || |||||||  | ||||
Db      22150 TACATTATACTGTAGTCTGTTAAGTGTGCAATAGCATTAGATCTTAAAAAAGTATATACT 22091

Qy       1170 TTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGT 1229
              ||||||| ||||| ||||||| |||||||||| ||||||||| ||||| |||||  ||
Db      22090 TTAATTTAAAAATACTTTATTGCTAAAAAATGCCAACAATCATCTGAGCCTTCAGCAAGA 22031

Qy       1230 TGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CGG 1278
              |||||||||||||||||| ||||||  |||  || |||  | |                ||
Db      22030 CATAATCTTTTTGCTGGTGGAGGGTCTTGCCTTGATGTTGATGGCTGCTAACTGATCAGG 21971

Qy       1279 GGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAA 1333
              |   | ||||||||  | | ||| ||||||| |||||||||      ||||||  |||
Db      21970 ATGATGGTTGCTGAAGGCTGGAGTGTCTGTGGCAATTTCTTAAAATAAGACAACAAGGAA 21911

Qy       1334 GATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTT 1393
              | |||| | |||| ||||||||||||||||||||||||||| ||| |||   || |||||
Db      21910 GTTTGCCACATCAATTGACTCTTCCTTTCATGAAAGATTTTTCTATAGCAAGTAATGCTG 21851

Qy       1394 TTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAAC 1452
              |||| |   || |||  |||||||||||||||||| ||| ||| |  ||||||||||||||
Db      21850 TTTGTTGAAATATTACCCACAGTAGAACTTCTTTCAAAGTTGAAGCCAATCCTCTCAAAC 21791

Qy       1453 CCTG-CTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTC 1511
              |||| | ||||||| |  | | ||||| |   |||||||| ||||| || ||||||||||
Db      21790 CCTGCCACTGCTTTTATCAACGAAGTTTACGGAATATTCTAAATCCTTTCTTGTCATTTC 21731

Qy       1512 AACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATC 1571
              ||||| ||||| ||  ||||| ||||| ||||||||||||||        || |    ||
Db      21730 AACAACTTTCATAGCATCTTCGCCAGGCGTAGATTCCATCTCAAGGAACCAGTT----TC 21675

Qy       1572 TTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCA 1631
              |||||||| |||| | ||| || ||||||||| ||||||||||||||||||||||||||||
Db      21674 TTTGCTCACCCATGAAAAGCAACTCCTCATCTCTTCAAGTTTTATCATGAGATTGCAGCA 21615

Qy       1632 ATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTAC 1691
              |  |||||  |||||| |      || || || |||||  |||||  |||||| |||| ||
Db      21614 GTTTAGTCACATCTTCAAG-----CTCCATTTCTAATTCTAGTTCACTTGCTATTTCCAC 21560

Qy       1692 CACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCCTCTCCAAGTCATCCATGAGGGT 1751
              |||||||   || ||||||||| |||||||||||||| ||| ||||||||||  ||||||
Db      21559 AACATCTGCAATTACTTCCTCCAGTGAAGTCTTGAACCCCTCAAAGTCATCCACGAGGGT 21500

Qy       1752 TGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATC 1808
              ||||||| || |||||||||||||||||| || || |||||||||   | |||  ||| ||
Db      21499 TGGAATCAACCTCTTCCAAATTCCTGTTCATGTTGATATTTTGATCTCTTCCTGTGAGTC 21440

Qy       1809 ATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACT 1868
              |||||||||||||||||| || | ||||||||| ||||||| |||||||||||||||||
Db      21439 CTGAATGTTCTTAATGGCATCTAGGATGGTGAATTCTTTCCAGAAGGTTTTCAATTTACT 21380

Qy       1869 TAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAAT 1928
              |   ||||| ||||||             | |  | ||| ||| ||||||||||| |||
Db      21379 TTACCCAGACCCATCC---------GAAGAATCACTATGTCAGCTATAGCCTTATGAAAT 21329

Qy       1929 GTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
              ||||||||| |||||||||  || |||||||||| ||||||||||
Db      21328 GTATTTCTTAAATAATAAGAGTTAAAAGTTGAAATGACTCCTTGATC 21282


RESULT 45
AL357520
LOCUS       AL357520              178942 bp    DNA     linear   PRI 18-MAY-2005
```

```
DEFINITION  Human DNA sequence from clone RP11-288D15 on chromosome 10 Contains
            a pseudogene similar to part of glycogen synthase kinase 3 and a
            ubiquitin-conjugating enzyme pseudogene, complete sequence.
ACCESSION   AL357520
VERSION     AL357520.12  GI:15620611
KEYWORDS    HTG; ubiquitin-conjugating.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 178942)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 14, 2001 this sequence version replaced gi:14485342.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-288D15 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .178942
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-288D15"
                     /clone_lib="RPCI-11.1"
     misc_feature    1
                     /note="Clone_left_end: RP11-288D15"
     misc_feature    43725. .43746
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    54175
                     /note="Clone_right_end: RP11-576H16"
     gene            66599. .67051
                     /locus_tag="RP11-288D15.1-001"
                     /pseudo
     CDS             66599. .67051
                     /locus_tag="RP11-288D15.1-001"
                     /note="match: proteins: AAF45801 AAH27984 AAM52705
                     AAN09082 AAN09083 P18265 P18431 P49840 Q8MRF7 Q91757
                     Q9IBD3 Q9NL43 Q9YH61"
                     /pseudo
                     /codon_start=1
                     /product="pseudogene similar to part of glycogen synthase
                     kinase 3"
     misc_feature    90945
                     /note="Clone_left_end: RP11-498A5"
     misc_feature    104374
                     /note="Clone_right_end: RP11-368F13"
     gene            join(177553. .177735,178054. .178268)
                     /locus_tag="RP11-288D15.2-001"
                     /pseudo
     CDS             join(177553. .177735,178054. .178268)
                     /locus_tag="RP11-288D15.2-001"
                     /note="match: proteins: Q13403 Q13404 Q15819 Q8K2V7 Q8VEB5
                     Q90879 Q96H34 Q9CYD7 Q9CZY3 Q9D2M8 Q9ERI8 Q9GZW1 Q9H4J4
                     Q9N1C3 Q9SVD7 Q9UM49 Q9VRL1"
                     /pseudo
                     /codon_start=1
                     /product="ubiquitin-conjugating enzyme pseudogene"
     misc_feature    178942
                     /note="Clone_right_end: RP11-288D15"
ORIGIN

  Query Match            25.2%;  Score 503.4;  DB 8;  Length 178942;
  Best Local Similarity  73.6%;  Pred. No. 2.2e-69;
  Matches  766;  Conservative    0;  Mismatches  231;  Indels   44;  Gaps    8;

Qy        974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
              || || ||| | ||| | |||||| ||||| ||  |||||| ||||||| |||||
Db      49673 TATTATAGGTAAACCATAGAGATATTGTGGATTCAGTTCCAGACCACCATAATAAG---- 49728

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
                |||||||||| ||| |||||||| | |  | | |||||| ||||||||
Db      49729 ----------GTGGATATCATAATTAAGTGAGTTAAAATAATTTTTTGGTTTCCCAGTA 49777

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
              | || ||||||| |  ||| |||| |   |||||| ||||||| ||||| | |||||||
Db      49778 CCTACAAAAGTTATATTTACACTATATCATAGTCTATTAAGTGCACAATAATATTATGTC 49837

Qy       1154 TAAAAAA-ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212
              ||||||| | | |||||||||||  ||||| |||||| ||||||| |||||| ||||||
Db      49838 TAAAAAATATAAATACCTTAATTAAAAAATACTTTATTGCTAAAAACTGCTAATAATCAT 49897

Qy       1213 TTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACT 1272
                |||  |||||||| ||||| || |||||| |  |||  ||||||  |||  || |||
Db      49898 CGGAGTCTTCAGTGAATTGTACTCCTTTTGCCGACGGAGGGTCTTGCCTTGATGTTGATG 49957

Qy       1273 GAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAA 1321
              |||           ||| |   | |||||||| || | |||||| ||| | |||||||||
Db      49958 GCTGCTGACTGATCAGGGTGGTGGTTGCTGAAGATTGGAGTGGCTCTGGAAATTTCTTAA 50017

Qy       1322 AACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAG 1381
              ||||||  ||| | |  | | ||   |||| |||||  || ||||||||||| ||| |||
Db      50018 GACAACATTGAGGTTCACCACATTGATTGATTCTTCACTTTATGAAAGATTTATCTATAG 50077

Qy       1382 TGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCA 1440
              |||| ||| | | | |||||||||||  ||||| ||||| || || ||||||||| |||
Db      50078 CGTGTAATGTTGTCTCATAGCATTTTACCCACAGGAGAACGTCCTTTAAAATTGGAGTCA 50137

Qy       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500
              ||| ||||| ||| |||  |   || |  | ||||||| || ||||| ||| || || ||
Db      50138 ATCTTCTCATACCATGCCATTGCTTTATCAACTAAGTTGATGTAATACTCTAAACCCTTT 50197

Qy       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560
              |||||||||||| ||| |||||||  ||||||| ||||||||||||||||||||
Db      50198 GTTGTCATTTCAATAATGTTCACAGCATCTTCACTAGGAGTAGATTCCATCTCA----AG 50253

Qy       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
              |  |    |||||||| || | ||||  | || | |||||  ||| ||||||| |||||
Db      50254 AAACTACTTTCTTTGCTTATTCGTAAGGGGCAACTACTCATTCGTTGAAGTTTTGTCATG 50313

Qy       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680
              |||||||||||| ||||||  | ||||| |   | |||   ||| | | |         |
Db      50314 AGATTGCAGCAATTCAGTCAGATTTTCAGTCTCTATTTCCAATTTTAGTTC------TCT 50367

Qy       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740
              | |||  |||||||||  |||  |||||||| ||||||||||||  ||| ||||||
Db      50368 CGTATTTTGACCACATCTGCAGTTGTTTCCTCCACTGAAGTCTTGAACTCCTCAAAGTCA 50427
```

```
Qy       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800
             ||||||||  || ||||  |||| |||||||| || |||||||| || |||||||| |||| |
Db      50428 TCCATGAGATTTAGAATAAACTTTTTCCAAACTCTTGTTAATGTTGATATTTTTACCTTC 50487

Qy       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860
             |||||||||||||||| ||||||||| || || ||||| ||||||| || |||||||||
Db      50488 CATGAATCATGAATGTTGTTAATGGCATCTAGACTGGTGTATCCTTTGCATAAGGTTTTC 50547

Qy       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCC- 1919
             || || |||   |||||| |||     |||||||  || |||  ||| ||| |||||||
Db      50548 AACTTTCTTTTCCCAGATTCAT-----CAGAGGAATCAATTTTTATGGCAGCTATAGCCT 50602

Qy       1920 TTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATT 1979
             ||||| |||  |||||||| ||||||||    |||||||   |||||||  |||||||||
Db      50603 TTATGAAATAGATTTCTTAAATAATAATTGATGAAAGTCAGAATTACTTTTTGATCCATG 50662

Qy       1980 TTCTGCAAAATAGATGTTGTG 2000
                |||||  || |||||||||
Db      50663 AGCTGCAGGATGGATGTTGTG 50683
```

Search completed: January 22, 2006, 13:37:04
Job time : 6837.28 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2006, 09:21:22 ; Search time 804.678 Seconds
(without alignments)
16564.866 Million cell updates/sec

Title: US-09-728-552A-3_COPY_1_2000
Perfect score: 2000
Sequence: 1 gaattctcctgcctcagcct.....tctgcaaaatagatgttgtg 2000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2000	100.0	80595	2	AAV83939	Aav83939 HC-contig
	2	1611.4	80.6	80240	2	AAV83940	Aav83940 NC-contig
c	3	636.8	31.8	2418	7	ADS31115	Ads31115 Human gen
c	4	636.8	31.8	2418	7	ADY36503	Ady36503 HIRA geno
	5	526.6	26.3	14033	6	ADH22218	Adh22218 Human cho
	6	516	25.8	3895	11	ADM01834	Adm01834 Human cDN
	7	505.2	25.3	88759	13	ABD32923_7	Continuation (8 of
	8	505.2	25.3	110000	13	ABD32923_6	Continuation (7 of
c	9	500.6	25.0	55829	13	ABD33512	Abd33512 Human can
	10	488	24.4	110000	12	ADQ97331_3	Continuation (4 of
c	11	480.2	24.0	21224	5	ABA20544	Aba20544 Human ner
c	12	480	24.0	32192	4	AAL04361	Aal04361 Human rep
c	13	479.6	24.0	198522	11	ACN44010	Acn44010 Human gen
c	14	477.6	23.9	2448	10	ADA53284	Ada53284 Human cod
c	15	475.8	23.8	276276	11	ACN44350	Acn44350 Human gen
	16	469.4	23.5	163350	6	AAD46127	Aad46127 Human tum
	17	466.6	23.3	110000	10	AAL52246_1	Continuation (2 of
	18	465.8	23.3	262090	12	ADQ59207	Adq59207 MSI-H car
	19	465.6	23.3	23271	13	ACN37213	Acn37213 Human per
	20	457.4	22.9	173564	13	ABD32953	Abd32953 Human can
c	21	454.8	22.7	105219	11	ACN44286	Acn44286 Human gen
	22	454.2	22.7	277616	13	ABD32602	Abd32602 Human can
	23	454	22.7	32190	4	AAK90524	Aak90524 Human dig
	24	453	22.7	96587	10	ADC85392	Adc85392 Human Jak
	25	453	22.7	96589	9	ADA02912	Ada02912 Human JAK
	26	453	22.7	96589	10	ADB72650	Adb72650 Human JAK
	27	453	22.7	96589	12	ADM74507	Adm74507 Human car
c	28	451.6	22.6	14117	6	ABL64107	Abl64107 Breast ca
	29	447.8	22.4	4503	4	AAL05284	Aal05284 Human rep
	30	446.8	22.3	175590	10	ADD50650	Add50650 BAC seque
	31	446.8	22.3	175590	14	ADV77908	Adv77908 Human BAC
	32	444	22.2	160771	6	ABQ88179	Abq88179 Human ost
	33	443.6	22.2	4320	4	AAK90584	Aak90584 Human dig
	34	443.6	22.2	4320	4	AAK90585	Aak90585 Human dig
c	35	442.6	22.1	32767	14	AEB68407	Aeb68407 Human glu
	36	441.6	22.1	110000	6	ABX08336_06	Continuation (7 of
	37	441.6	22.1	110000	12	ADJ25985_06	Continuation (7 of
	38	441.6	22.1	110000	12	ADN97989_06	Continuation (7 of
	39	441.6	22.1	110000	12	ADO50281_06	Continuation (7 of
	40	441.6	22.1	110000	14	AEB85185_06	Continuation (7 of
c	41	441	22.1	96589	9	ADA02954	Ada02954 Human NR3
c	42	441	22.1	96589	10	ADB72692	Adb72692 Human NR3
c	43	441	22.1	96589	10	ADC85434	Adc85434 Human Nr3
c	44	441	22.1	96589	12	ADM74549	Adm74549 Human car
	45	439.8	22.0	128779	12	ADF13116	Adf13116 Hypermeth
	46	439.8	22.0	128779	12	ADI37262	Adi37262 Hypermeth
c	47	439.2	22.0	102457	9	ACH03408_3	Continuation (4 of
c	48	438.4	21.9	6022	5	ABA18218	Aba18218 Human ner
	49	437	21.9	175077	11	ACN44626	Acn44626 Human gen
	50	436.8	21.8	182328	12	ADL08128	Adl08128 Human gen
	51	435.8	21.8	110000	13	ABD32911_6	Continuation (7 of
c	52	434.8	21.7	125439	6	ABQ88177	Abq88177 Human ost
	53	432.4	21.6	209484	11	ACN44126	Acn44126 Human gen
c	54	429.2	21.5	4015	13	ADR06850	Adr06850 Full leng
	55	428.4	21.4	337022	12	ADQ59416	Adq59416 Human can
	56	428.4	21.4	338780	14	ADZ13691	Adz13691 Human can
	57	428.2	21.4	221600	14	ADX80720	Adx80720 Human neu
c	58	425.6	21.3	110000	4	AAK95240_04	Continuation (5 of
c	59	425.6	21.3	110000	4	AAK96733_04	Continuation (5 of
c	60	425.6	21.3	110000	6	ABT00010_04	Continuation (5 of
c	61	425.6	21.3	110000	6	ABT01503_04	Continuation (5 of
c	62	425.6	21.3	110000	11	ADW70291_04	Continuation (5 of
c	63	425.6	21.3	110000	12	ADH77486_04	Continuation (5 of
c	64	419.6	21.0	160203	10	ADL13732	Adl13732 Osteoarth
	65	418.8	20.9	30265	13	ACN37210	Acn37210 Human per
	66	418.8	20.9	30265	14	ADZ59516	Adz59516 Secondary
	67	418.8	20.9	87320	14	ADZ12863	Adz12863 Human can
c	68	416	20.8	112132	6	ABK90888	Abk90888 Human ATP
c	69	416	20.8	112132	11	ADM56267	Adm56267 Human ATP
c	70	416	20.8	112132	14	ADV85874	Adv85874 Human pro
	71	415.8	20.8	220860	12	ADN36595	Adn36595 Human pro
	72	414.4	20.7	121600	6	ABT10748	Abt10748 Human bre
	73	414.4	20.7	121600	12	ADQ18307	Adq18307 Human sof
c	74	413.4	20.7	79860	11	ACN44122	Acn44122 Human gen
c	75	412.2	20.6	52691	4	AAK84435	Aak84435 Human imm
c	76	411.2	20.6	236303	4	AAS11614	Aas11614 Human gen
c	77	410.6	20.5	301477	13	ABD33362	Abd33362 Human can
c	78	410.4	20.5	72215	4	AAK86832	Aak86832 Human imm
c	79	410.4	20.5	173805	10	ADL13775	Adl13775 Osteoarth
c	80	409.6	20.5	110000	12	ADO79173_2	Continuation (3 of
	81	408.4	20.4	110000	9	ADA13411_1	Continuation (2 of
	82	408.4	20.4	110000	12	ADQ59446_1	Continuation (2 of
	83	408.4	20.4	110000	14	ADZ13757_1	Continuation (2 of
	84	407.8	20.4	110000	6	ABX08336_01	Continuation (2 of
	85	407.8	20.4	110000	12	ADJ25985_01	Continuation (2 of
	86	407.8	20.4	110000	12	ADN97989_01	Continuation (2 of
	87	407.8	20.4	110000	12	ADO50281_01	Continuation (2 of
	88	407.8	20.4	110000	14	AEB85185_01	Continuation (2 of
c	89	405.8	20.3	83946	6	ABQ88101	Abq88101 Human ost
c	90	405.8	20.3	83946	12	ADQ17978	Adq17978 Human sof
c	91	405.4	20.3	2650	6	ABV75666	Abv75666 Human FD1
c	92	404.2	20.2	349989	10	ADC86916	Adc86916 Human GPC

c	93	403.4	20.2	15809	4	AAK68321
	94	401	20.1	106544	13	ABD32669
	95	399	20.0	76600	12	ADO79402
	96	396.2	19.8	101569	12	ADP13274
	97	396.2	19.8	147309	6	ABK49450
c	98	394.4	19.7	117813	14	ADX07540
c	99	393.8	19.7	110000	10	ADE11169_1
	100	393	19.7	112190	4	AAH44801

Aak68321 Human imm
Abd32669 Human can
Ado79402 DLG1 regi
Adp13274 Renal cel
Abk49450 Human tra
Adx07540 Cyclin-de
Continuation (2 of
Aah44801 Human GPC

ALIGNMENTS

RESULT 1
AAV83939
ID AAV83939 standard; DNA; 80595 BP.
XX
AC AAV83939;
XX
DT 03-MAR-1999 (first entry)
XX
DE HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 8; Fig 6; 540pp; English.
XX
CC The present sequence represents the HC-contig derived from normal human
CC chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC produce an unusual chromosomal marker referred to as mardel(10). The
CC mardel(10) marker is mitotically stable and contains a functional
CC neocentromere at a location regarded as non-centromeric. This
CC neocentromere maps to q25.2 on chromosome 10. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in a
CC compatible cell, of replicating, acting as extra-chromosomal element and
CC segregating during cell division. The sequences can be used to construct
CC artificial chromosomes for use in gene therapy comprising a replicable,
CC segregating nucleic acid that confers a specific phenotype on cells.
CC Human artificial chromosomes can propagate in human cells and carry large
CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC they are not mutagenic. The artificial chromosomes are also useful for
CC generation of transgenic plants and animals, in production of proteins
CC and to make diagnostic reagents, e.g. for expression of cytokines,
CC receptors and growth factors, or to increase the copy number of a gene in
CC a cell. The constructs may also be used for functional and structural
CC analysis of chromosomes
XX

SQ Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

Query Match 100.0%; Score 2000; DB 2; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60
Db       1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60

Qy      61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120
Db      61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120

Qy     121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180
Db     121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180

Qy     181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240
Db     181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240

Qy     241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300
Db     241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300

Qy     301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 360
Db     301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAATATTTTTCTG 360

Qy     361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420
Db     361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420

Qy     421 TATTTTTCCTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480
Db     421 TATTTTTCCTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480

Qy     481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540
Db     481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540

Qy     541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600
Db     541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600

Qy     601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660
Db     601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660

Qy     661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720
Db     661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720

Qy     721 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780
Db     721 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780

Qy     781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840
Db     781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840

Qy     841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATGTCTATAATACATT 900
Db     841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATGTCTATAATACATT 900

Qy     901 AGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCT 960
Db     901 AGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCT 960

Qy     961 GTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCAC 1020
Db     961 GTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCAC 1020
```

```
Qy       1021 CACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTT 1080

Qy       1081 TGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA 1140

Qy       1141 ATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAAT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAAT 1200

Qy       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260

Qy       1261 TTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320

Qy       1321 AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCA 1440

Qy       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500

Qy       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560

Qy       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620

Qy       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680

Qy       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740

Qy       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800

Qy       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860

Qy       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920

Qy       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980

Qy       1981 TCTGCAAAATAGATGTTGTG 2000
             ||||||||||||||||||||
Db       1981 TCTGCAAAATAGATGTTGTG 2000
```

RESULT 2
AAV83940
ID AAV83940 standard; DNA; 80240 BP.
XX
AC AAV83940;
XX
DT 03-MAR-1999 (first entry)
XX
DE NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 9; Fig 16A; 540pp; English.
XX
CC The present sequence represents the NC-contig derived from a mutated
CC human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC mitotically stable and contains a functional neocentromere at a location
CC regarded as non-centromeric. This neocentromere maps to q25.2 on
CC chromosome 10. The specification describes nucleic acid sequences derived
CC from a eukaryotic chromosome, including a neocentromere or its functional
CC derivative or hybrid, that are able, in a compatible cell, of
CC replicating, acting as extra-chromosomal element and segregating during
CC cell division. The sequences can be used to construct artificial
CC chromosomes for use in gene therapy comprising a replicable, segregating
CC nucleic acid that confers a specific phenotype on cells. Human artificial
CC chromosomes can propagate in human cells and carry large amounts of DNA
CC (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC mutagenic. The artificial chromosomes are also useful for generation of
CC transgenic plants and animals, in production of proteins and to make
CC diagnostic reagents, e.g. for expression of cytokines, receptors and
CC growth factors, or to increase the copy number of a gene in a cell. The
CC constructs may also be used for functional and structural analysis of
CC chromosomes
XX
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

Query Match 80.6%; Score 1611.4; DB 2; Length 80240;
Best Local Similarity 97.4%; Pred. No. 2.4e-264;
Matches 1666; Conservative 2; Mismatches 40; Indels 3; Gaps 3;

```
Qy        292 GATTGATCTGTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAAT 351
              | ||||| |||   || | | || || | ||| |||| ||  | |||| |||||  |||
Db          2 GGTTGATTTGTNNATAAGGAAGTTTGGAATCAATCCCGGAAGGAATTTTTTTTTTAAAAA 61

Qy        352 ATTTTTCTGAATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTT 411
              ||||||  ||| | || || ||:||| | |:||| | ||| ||||||||| ||| | |||
Db         62 ATTTTTTGGAAGG-GTTTGGTAWTAAAAAARCCAATTTGGGTTTTTAAAAATAGGAATTT 120

Qy        412 TATGGTATATATTTTTCCTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATAC 471
              ||||| | | | |||||| ||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 TATGGGAAAAAATTTTCCCTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATAC 180
Qy        472 TTAAAGTGGGTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGAT 531
Db        181 TTAAAGTGGGTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGAT 240
Qy        532 AATCTCAACCTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATG 591
Db        241 AATCTCAACCTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATG 300
Qy        592 GTTTGATTTGCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTC 651
Db        301 GTTTGATTTGCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTC 360
Qy        652 TTTTTTCATCTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATC 711
Db        361 TTTTTTCATCTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATC 420
Qy        712 TATTGGCTTTTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAA 771
Db        421 TATTGGCTTTTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAA 480
Qy        772 TATACATCTTTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAA 831
Db        481 TATACATCTTTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAA 540
Qy        832 TGTAGAAACCTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTA-TGTCT 890
Db        541 TGTAGAAACCTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATTGTCT 600
Qy        891 ATAATACATTAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCA 950
Db        601 ATAATACATTAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCA 660
Qy        951 AACATTTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGT 1010
Db        661 AACATTTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGT 720
Qy       1011 TCCATACCACCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCAC 1070
Db        721 TCCATACCACCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCAC 780
Qy       1071 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 1130
Db        781 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 840
Qy       1131 TAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATT 1190
Db        841 TCAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATT 900
Qy       1191 ACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGA 1250
Db        901 ACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGA 960
Qy       1251 AGGTCTTTTCTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGG 1310
Db        961 AGGTCTTTTCTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGG 1020
Qy       1311 CAGTTTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGA 1370
Db       1021 CAGTTTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGA 1080
Qy       1371 TTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAA 1430
Db       1081 TTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAA 1140
Qy       1431 AATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATT 1489
Db       1141 AATTGGAGTCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATT 1200
Qy       1490 CTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCA 1549
Db       1201 CTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCA 1260
```

```
Qy       1550 TCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAA 1609
Db       1261 TCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAA 1320
Qy       1610 GTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATT 1669
Db       1321 GTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATT 1380
Qy       1670 CCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACC 1729
Db       1381 CCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACC 1440
Qy       1730 TCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATA 1789
Db       1441 TCTCCAAGTCATCCATGAGGGCTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATA 1500
Qy       1790 TTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCC 1849
Db       1501 TTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCC 1560
Qy       1850 AAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCC 1909
Db       1561 AAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCC 1620
Qy       1910 AGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCC 1969
Db       1621 AGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCC 1680
Qy       1970 TTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
Db       1681 TTGATCCATTTTCTGCAAAATAGATGTTGTG 1711
```

RESULT 3
ADS31115/c
ID ADS31115 standard; DNA; 2418 BP.
XX
AC ADS31115;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genome high complexity repeat found in the HIRA gene #148.
XX
KW Human; ds;
KW histone cell cycle regulation defective, S. cerevisiae homologue A; HIRA;
KW high complexity repeat; in situ hybridisation; Southern blot;
KW chromosome breakpoint; inherited genetic disease; neoplastic disorder;
KW chromosome 22; DiGeorge syndrome; Velo-Cardio-facial syndrome.
XX
OS Homo sapiens.
XX
PN US2003224356-A1.
XX
PD 04-DEC-2003.
XX
PF 14-MAY-2001; 2001US-00854867.
XX
PR 16-MAY-2000; 2000US-00573080.
XX
PA (KNOL/) KNOLL J H M.
PA (ROGA/) ROGAN P K.
XX
PI Knoll JHM, Rogan PK;
XX
DR WPI; 2002-062378/08.
XX
PT Single copy genomic hybridization probes for detecting specific nucleic
PT acid sequences in sample by in situ hybridization useful for detection of
PT acquired or inherited genetic diseases.
XX
PS Example 1; SEQ ID NO 148; 30pp; English.
XX

```
CC   The invention relates to a nucleic acid hybridisation probe comprising a
CC   labelled, single copy nucleic acids of at least 50 nucleotides, which
CC   will hybridise to a deduced single copy sequence interval in target
CC   nucleic acid (TNA) of known sequence. The single copy sequence is deduced
CC   by comparing the target nucleic acid (e.g. a disease causing gene) with a
CC   collection of high and low complexity repeat sequences as found in the
CC   genome of the organism from containing the target nucleic acid. The probe
CC   is generated by PCR on the target sequence. The probe is essentially free
CC   of blocking nucleic acid sequences which will hybridise to repeat
CC   sequences within the genome of which the TNA is a part, and is labelled
CC   with a label selected from fluorochrome-responsive labels, fluorochromes,
CC   calorimetric chemical, conjugated proteins, antibodies, antigens and
CC   their mixtures. The probe is useful in a hybridisation method, where the
CC   hybridisation method is from in situ hybridisation, Southern blot, and
CC   other methods in which nucleic acid is immobilised, where the method
CC   further comprises selecting a single copy nucleic acid which will
CC   hybridise to a duplicon or triplicon sequence domain. The probe is useful
CC   for determining the existence of previously unknown repeat sequence
CC   families in a genome. The method comprises reacting a labelled probe with
CC   the genome, causing the probe to hybridise and ascertaining if the probe
CC   hybridises to the genome at more than three preferably ten different
CC   locations as a determination of new repeat sequence family, where the
CC   determining step comprises selecting the single copy sequence from a
CC   duplicon or triplicon sequence domain. The probe is useful for
CC   determining a chromosome breakpoint and is useful in the fields for
CC   cytogenetics and molecular genetics for determining the presence of
CC   specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
CC   the probes may be used to analyse specific chromosomal locations by in
CC   situ hybridisation as a detection of acquired or inherited genetic
CC   diseases especially for detection of genetic or neoplastic disorders.
CC   Unlike prior art techniques, the probe permits more precise chromosomal
CC   breakpoint determinations by in situ hybridisation. The genomic sequence
CC   comprising the human HIRA gene (histone cell cycle regulation defective,
CC   S. cerevisiae, homologue A) was analysed for single copy sequence
CC   intervals for use as probes of the invention. HIRA is located on
CC   chromosome 22 as a duplicate, deletions of 1 copy lead to DiGeorge and
CC   Velo-Cardio-facial syndromes. The present sequence is a high complexity
CC   repeat found within the human genome used to analyse the HIRA gene for
CC   repeat regions. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030224356.
XX
SQ   Sequence 2418 BP; 753 A; 472 C; 517 G; 668 T; 0 U; 8 Other;

  Query Match             31.8%;  Score 636.8;  DB 7;  Length 2418;
  Best Local Similarity   82.1%;  Pred. No. 5.5e-99;
  Matches  859;  Conservative    1;  Mismatches  133;  Indels   53;  Gaps    9;

Qy        979 CAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATG 1038
              |||||||||| ||||||| || ||||| ||||||| |||||| |||||
Db       2418 CAGGCATACCTCGGAGATATTGCGGGTTCGGTTCCAGACCACCGCAATA---------- 2370

Qy       1039 CAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATAT 1098
                 ||| | ||||| |||||||| ||||||||| | | |||||| ||||||||||||||
Db       2369 ----AAGCGAATATCGCAATAAAGCGAGTCACACGAATTTTTTGGTTTCCCAGTGCATAT 2314

Qy       1099 AAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAA 1158
              ||||||| || |||  ||| |||||||||| |||||||||||||||  ||||||||||||
Db       2313 AAAAGTTATGTTTACGCTATACTGTAGTCTATTAAGTGTGCAATAGCATTATGTCTAAAA 2254

Qy       1159 AA----ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTT 1214
              ||      |||||||||||||| ||||| ||||||| |||||||||||||| ||||| |
Db       2253 AACAATGTACATACCTTAATTTAAAAATACTTTATTGCTAAAAAATGCTAACGATCATCT 2194

Qy       1215 GAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGA 1274
              |||| ||||| |||| |||||||||||||||||||| ||||||   |||  || ||| |
Db       2193 GAGCCTTCAGCGAGTCGTAATCTTTTTGCTGGTGGAGGGTCTTGCCTTGATGTTGATGGC 2134

Qy       1275 T-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA- 1322
              |           ||| |   | |||||||| || |||||||||||||| |||||||||
Db       2133 TGCTGACTGATCAGGGTGGTGGTTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCTTAAAA 2074

Qy       1323 ----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTC 1378
                  |||||| ||||| ||||   |||| ||||||||||||||||||||||||||||||
Db       2073 TAAGACAACAATGAAGTTTGCCGCATCAATTGACTCTTCCTTTCATGAAAGATTTCTCTG 2014

Qy       1379 TAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA- 1437
              |||  |||||||||| ||||| |||||||||  |||||||||||||||| |||||||||
Db       2013 TAGCATGTGATGCTGTTTGACAGCATTTTACCCACAGTAGAACTTCTTTCAAAATTGGAG 1954

Qy       1438 TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCC 1497
              ||||||||||||||||||||       || |  | |||||||  || |||||||| |||||
Db       1953 TCAATCCTCTCAAACCCTGCCGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCC 1894

Qy       1498 ATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTT 1557
              |||||||||||||||||| |||||||  |||||||||||||||||||||||||||||||
Db       1893 TTTGTTGTCATTTCAACAATGTTCACAGCATCTTCACCAGGAGTAGATTCCATCTCAAGA 1834

Qy       1558 CCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATC 1617
                 | |    |||||||||||||||||||| || |||||||| ||||||||||||||
Db       1833 AACCACTT----TCTTTGCTCATCCATAAGAAGCAACTCCTCATCCGTTCAAGTTTTATC 1778

Qy       1618 ATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCT 1677
              |||||||||||||||| ||||||  ||||||||      || |||||  |||||| ||||||
Db       1777 ATGAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCCACTTCTAATTCTAGTTCT 1723

Qy       1678 CTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAG 1737
              |||||| |||| |||||||||||  ||| |||||||||:|||||||||:|||| ||| |||
Db       1722 CTTGCTATTTCCACCACATCTGCAGTTACTTCCTCCACTGAAGTCTTRAACCCCTCAAAG 1663

Qy       1738 TCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA-- 1795
              |||||||||||||||||||| |||||||||||| |||||||||| || |||||||||
Db       1662 TCATCCATGAGGGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGATATTTTGACC 1603

Qy       1796 -CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAG 1854
              |||||||||||||| |||||||||||||||||| || |||||||||||||||||| |||
Db       1602 TCCTCCCATGAATCACGAATGTTCTTAATGGCATCTAGAATGGTGAATCCTTTCCAGAAG 1543

Qy       1855 GTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTA 1914
              |||||||||||||| | |||||||||      |||||||  |||| || ||| ||| ||
Db       1542 GTTTTCAATTTACTTTGCCCAGATCCAT-----CAGAGGAATCACTATCTATGGCAGCTA 1488

Qy       1915 TAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGAT 1974
              |||||||||  |||||||||||| |||||||| |||||||| ||||||||||||||||
Db       1487 TAGCCTTATAAAATGTATTTCTTAAATAATAAGACTTGAAAGTCGAAATTACTCCTTGAT 1428

Qy       1975 CCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||   ||||| ||| |||||||||
Db       1427 CCATGGGCTGCAGAATGGATGTTGTG 1402


RESULT 4
ADY36503/c
ID   ADY36503 standard; DNA; 2418 BP.
XX
AC   ADY36503;
XX
DT   05-MAY-2005  (first entry)
XX
DE   HIRA genomic fragment SEQ ID NO 148.
XX
KW   hybridization; DNA detection; neoplasm; genetic disorder; cytogenetics;
KW   HIRA; ds.
XX
OS   Homo sapiens.
XX
PN   WO200188089-A2.
XX
PD   22-NOV-2001.
XX
PF   15-MAY-2001; 2001WO-US015674.
XX
PR   16-MAY-2000; 2000US-00573080.
PR   14-MAY-2001; 2001US-00854867.
XX
PA   (CHIL-) CHILDREN'S MERCY HOSPITAL.
```

XX
PI Knoll JHM, Rogan PK, Cazarro PM;
XX
DR WPI; 2002-062378/08.
XX
PT Single copy genomic hybridization probes for detecting specific nucleic
PT acid sequences in sample by in situ hybridization useful for detection of
PT acquired or inherited genetic diseases.
XX
PS Example 1; SEQ ID NO 148; 67pp; English.
XX
CC The invention describes a nucleic acid hybridization probe (I) comprising
CC a labeled, single copy nucleic acids of at least 50 nucleotides, which
CC will hybridize to a deduced single copy sequence interval in target
CC nucleic acid (TNA) of known sequence. (I) is useful in a hybridization
CC method which comprises preparing a reaction mixture comprising TNA and
CC (I) which hybridizes to TNA, and causing (I) to hybridize to TNA, where
CC the hybridization method is from in situ hybridization, Southern blot,
CC and other methods in which nucleic acid is immobilized, where the method
CC further comprises selecting a single copy nucleic acid which will
CC hybridize to a duplicon or triplicon sequence domain. (I) is useful for:
CC determining the existence of previously unknown repeat sequence families
CC in a genome; determining a chromosome breakpoint and in the fields of
CC cytogenetics and molecular genetics for determining the presence of
CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
CC the probes may be used to analyze specific chromosomal locations by in
CC situ hybridization as a detection of acquired or inherited genetic
CC diseases especially for detection of genetic or neoplastic disorders.
CC Unlike prior art techniques, (I) permits more precise chromosomal
CC breakpoint determinations by in situ hybridization. Hybridization
CC techniques utilizing (I), have made it possible to obtain reliable,
CC easily detectable signals with relatively small probes. A readily
CC detectable signal was obtained with a probe on the order of 2 kb in
CC length, using fluorescent in situ hybridization (FISH) technology. This
CC sensitivity of (I) is improved compared to the prior art, because the
CC probes of (I) are homogeneous single copy sequences. However, smaller
CC amplified segments, each comprising non-repetitive sequences, may also be
CC used in combination as probes to achieve adequate signals for in situ
CC hybridization. Complex single copy probes that hybridize to duplicated or
CC triplicated targets can also increase hybridization signals. This
CC sequence represents a human HIRA genomic sequence that shows homology to
CC a known high-complexity repeat sequence family of the human genome and is
CC used in the creation of an HIRA gene probe.
XX
SQ Sequence 2418 BP; 753 A; 472 C; 517 G; 668 T; 0 U; 8 Other;

Query Match 31.8%; Score 636.8; DB 7; Length 2418;
Best Local Similarity 82.1%; Pred. No. 5.5e-99;
Matches 859; Conservative 1; Mismatches 133; Indels 53; Gaps 9;

```
Qy        979 CAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATG 1038
              |||||||||||| ||||||| || ||||| ||||||| |||||| |||||
Db       2418 CAGGCATACCTCGGAGATATTGCGGGTTCGGTTCCAGACCACCGCAATA---------- 2370

Qy       1039 CAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATAT 1098
                 ||| | ||||| |||||||| |||||||||| | | |||||| ||||||||||||||
Db       2369 ----AAGCGAATATCGCAATAAAGCGAGTCACACGAATTTTTTGGTTTCCCAGTGCATAT 2314

Qy       1099 AAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAA 1158
              ||||||| || |||  ||| |||||||||| ||||||||||||||  ||||||||||||
Db       2313 AAAAGTTATGTTTACGCTATACTGTAGTCTATTAAGTGTGCAATAGCATTATGTCTAAAA 2254

Qy       1159 AA----ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTT 1214
              ||      |||||||||||||| ||||| ||||||| |||||||||||||| ||||| |
Db       2253 AACAATGTACATACCTTAATTTAAAAATACTTTATTGCTAAAAAATGCTAACGATCATCT 2194

Qy       1215 GAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGA 1274
              |||| ||||| |||| |||||||||||||||||||| |||||| |||  || |||  |
Db       2193 GAGCCTTCAGCGAGTCGTAATCTTTTTGCTGGTGGAGGGTCTTGCCTTGATGTTGATGGC 2134

Qy       1275 T-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA- 1322
              |           ||| |   | |||||||| || |||||||||||||| |||||||||
Db       2133 TGCTGACTGATCAGGGTGGTGGTTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCTTAAAA 2074

Qy       1323 ----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTC 1378
                  |||||| ||||| ||||   |||| ||||||||||||||||||||||||||||||
Db       2073 TAAGACAACAATGAAGTTTGCCGCATCAATTGACTCTTCCTTTCATGAAAGATTTCTCTG 2014

Qy       1379 TAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA- 1437
              |||  |||||||| |||| |||||||||  ||||||||||||||||| |||||||||
Db       2013 TAGCATGTGATGCTGTTTGACAGCATTTTACCCACAGTAGAACTTCTTTCAAAATTGGAG 1954

Qy       1438 TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCC 1497
              |||||||||||||||||||       || |  | ||||||| || ||||||||| |||||
Db       1953 TCAATCCTCTCAAACCCTGCCGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCC 1894

Qy       1498 ATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTT 1557
              |||||||||||||||||| |||||||  ||||||||||||||||||||||||||||||
Db       1893 TTTGTTGTCATTTCAACAATGTTCACAGCATCTTCACCAGGAGTAGATTCCATCTCAAGA 1834

Qy       1558 CCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATC 1617
              |  |    |||||||||||||||||||||| || |||||||| ||||||||||||||||
Db       1833 AACCACTT----TCTTTGCTCATCCATAAGAAGCAACTCCTCATCCGTTCAAGTTTTATC 1778

Qy       1618 ATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCT 1677
              |||||||||||||||| ||||||  ||||||||      || ||||| |||||| ||||||
Db       1777 ATGAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCCACTTCTAATTCTAGTTCT 1723

Qy       1678 CTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAG 1737
              |||||| |||| |||||||||||  ||| |||||||||  |||||||||:|||| ||| |||
Db       1722 CTTGCTATTTCCACCACATCTGCAGTTACTTCCTCCACTGAAGTCTTRAACCCCTCAAAG 1663

Qy       1738 TCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA-- 1795
              |||||||||||||||||||| |||||||||||| |||||||||| || |||||||||
Db       1662 TCATCCATGAGGGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGATATTTTGACC 1603

Qy       1796 -CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAG 1854
              |||||||||||||| |||||||||||||||| || |||||||||||||||||| |||
Db       1602 TCCTCCCATGAATCACGAATGTTCTTAATGGCATCTAGAATGGTGAATCCTTTCCAGAAG 1543

Qy       1855 GTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTA 1914
              |||||||||||||| | ||||||||||     |||||||  |||| || ||| ||| ||
Db       1542 GTTTTCAATTTACTTTGCCCAGATCCAT-----CAGAGGAATCACTATCTATGGCAGCTA 1488

Qy       1915 TAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGAT 1974
              |||||||| |||||||||||| |||||||| |||||||| ||||||||||||||||||||
Db       1487 TAGCCTTATAAAATGTATTTCTTAAATAATAAGACTTGAAAGTCGAAATTACTCCTTGAT 1428

Qy       1975 CCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||  ||||| ||| |||||||||
Db       1427 CCATGGGCTGCAGAATGGATGTTGTG 1402
```

RESULT 5
ADH22218
ID ADH22218 standard; DNA; 14033 BP.
XX
AC ADH22218;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human cholinergic receptor, nicotinic, gamma (CHRNG) genomic DNA SeqID 1.
XX
KW human; ds; chromosome 2q33-q34; cholinergic receptor, nicotinic, gamma;
KW CHRNG; haplotype; drug discovery; acetylcholine receptor; AChR;
KW myasthenia gravis; screening assay; single nucleotide polymorphism; SNP;
KW gene.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	variation	replace(3697,t)
FT		/*tag= a
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(3883,t)
FT		/*tag= b

```
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3902,g)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             4066. .10034
FT                   /*tag=  e
FT                   /product= "CHRNG protein"
FT                   /note= "This coding sequence contains 12 exons"
FT   exon            4066. .4120
FT                   /*tag=  d
FT                   /number= 1
FT   intron          4121. .4309
FT                   /*tag=  f
FT                   /number= 1
FT   variation       replace(4198,c)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            4310. .4449
FT                   /*tag=  h
FT                   /number= 2
FT   intron          4450. .4698
FT                   /*tag=  i
FT                   /number= 2
FT   exon            4699. .4743
FT                   /*tag=  j
FT                   /number= 3
FT   intron          4744. .4919
FT                   /*tag=  k
FT                   /number= 3
FT   exon            4920. .5029
FT                   /*tag=  l
FT                   /number= 4
FT   intron          5030. .5691
FT                   /*tag=  m
FT                   /number= 4
FT   exon            5692. .5847
FT                   /*tag=  n
FT                   /number= 5
FT   intron          5848. .6740
FT                   /*tag=  o
FT                   /number= 5
FT   variation       replace(6606,g)
FT                   /*tag=  p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6728,t)
FT                   /*tag=  q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            6741. .6838
FT                   /*tag=  r
FT                   /number= 6
FT   intron          6839. .7199
FT                   /*tag=  s
FT                   /number= 6
FT   variation       replace(7124,c)
FT                   /*tag=  t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            7200. .7400
FT                   /*tag=  u
FT                   /number= 7
FT   variation       replace(7257,g)
FT                   /*tag=  v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7258,a)
FT                   /*tag=  w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          7401. .7592
FT                   /*tag=  x
FT                   /number= 7
FT   exon            7593. .7707
FT                   /*tag=  y
FT                   /number= 8
FT   intron          7708. .7902
FT                   /*tag=  z
FT                   /number= 8
FT   variation       replace(7847,g)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7891,t)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            7903. .8017
FT                   /*tag=  ac
FT                   /number= 9
FT   intron          8018. .8684
FT                   /*tag=  ad
FT                   /number= 9
FT   variation       replace(8042,c)
FT                   /*tag=  ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8621,a)
FT                   /*tag=  af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            8685. .8898
FT                   /*tag=  ag
FT                   /number= 10
FT   intron          8899. .9089
FT                   /*tag=  ah
FT                   /number= 10
FT   exon            9090. .9220
FT                   /*tag=  ai
FT                   /number= 11
FT   intron          9221. .9860
FT                   /*tag=  aj
FT                   /number= 11
FT   variation       replace(9337,a)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            9861. .10034
FT                   /*tag=  al
FT                   /number= 12
FT   variation       replace(9902,t)
FT                   /*tag=  am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10111,c)
FT                   /*tag=  an
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   WO200222643-A1.
XX
PD   21-MAR-2002.
XX
PF   17-SEP-2001; 2001WO-US029206.
XX
PR   15-SEP-2000; 2000US-0232807P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Gilson CR,  Koshy B,  Kliem SE,  Sausker EA;
XX
DR   WPI; 2002-371968/40.
DR   P-PSDB; ADH22134.
XX
PT   New genetic variants of cholinergic receptor, nicotinic, gamma
PT   polypeptide, CHRNG gene useful for therapeutic purposes and for
PT   expressing CHRNG protein useful in identifying drugs to treat myasthenia
PT   gravis.
XX
PS   Claim 1; SEQ ID NO 1; 107pp; English.
XX
CC   This invention relates to novel genetic markers and variants of the gene
CC   encoding the cholinergic receptor, nicotinic, gamma polypeptide (CHRNG),
CC   located on chromosome 2q33-p34. Specifically, it refers to a set of
CC   haplotypes in the CHRNG gene, which are useful for improving the
CC   efficiency and output of the drug discovery process by the identification
CC   of drugs that can target the CHRNG protein and treat disorders associated
CC   with its abnormal expression or function. The CHRNG protein is the gamma
```

CC subunit of the acetylcholine receptor (AChR), and autoantibodies directed
CC against the embryonic form of AChR play an important role in the
CC pathogenesis of neonatal myasthenia gravis. As such, the present
CC invention describes a method for identifying an association between a
CC trait (such as a clinical response to a drug that targets CHRNG) and a
CC haplotype or haplotype pair of the CHRNG gene. Furthermore, it is useful
CC in screening assays, for the development of diagnostic tests and for
CC therapeutic treatments of myasthenia gravis using gene therapy. This
CC polynucleotide is the human CHRNG genomic DNA sequence that contains the
CC SNPs of the invention.
XX
SQ Sequence 14033 BP; 3341 A; 3737 C; 3747 G; 3192 T; 0 U; 16 Other;

Query Match 26.3%; Score 526.6; DB 6; Length 14033;
Best Local Similarity 78.7%; Pred. No. 3e-80;
Matches 767; Conservative 0; Mismatches 169; Indels 39; Gaps 10;

```
Qy       1051 ATCACAATAAAGTGAGTCACA-CAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGC 1109
Db      12263 AACACAAGAAAGCAAGTCACATGAATGTTTTTGGTTTCCCAATGCATATAAAAGTTATGT 12322

Qy       1110 TTATACTACACTGTAGTCTGTTAAGTGTGCAATAG---TGTTATGTCTAAAAAAACACAT 1166
Db      12323 TTACACTATACTGTAGACCAGCAAGAGTGCAATAGCATTGTCTAATAAAACAATATACAT 12382

Qy       1167 ACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC-ATTTGAGCATTCAGT 1225
Db      12383 ACCTAAATTTAAAAATACTTTATTGCTAAAAAATGCTGATTATCAATCTGAGCCTTCGGT 12442

Qy       1226 GAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---------- 1275
Db      12443 GAGTCGTACTCTTTTTGCTGGTGGAGGGTCTTGCCTCGAGGTTGATGGTTGCTGACTGGT 12502

Qy       1276 -CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAG 1329
Db      12503 CAGGGTGGTGGCTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCCTAAAATAAGACATCGG 12562

Qy       1330 TGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGAT 1389
Db      12563 TGAAGCTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTGTAGCATGCGAT 12622

Qy       1390 GCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTC 1448
Db      12623 GCTGTTTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAGTCAATCCTCTC 12682

Qy       1449 AAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCAT 1508
Db      12683 AAACCCTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGTTGTCAT 12742

Qy       1509 TTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGA 1568
Db      12743 TTCAACAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGAAACCACTT--- 12799

Qy       1569 ATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCA 1628
Db      12800 -TCTTTGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATCATGAGATTGCA 12858

Qy       1629 GCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTC 1688
Db      12859 GCAATTCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCTTTTGCTATTTC 12913

Qy       1689 TACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAG 1748
Db      12914 CACCACATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAGTCATCAATGAG 12973

Qy       1749 GGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGA 1805
Db      12974 GGTGGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACCTCCTCCCATGA 13033

Qy       1806 ATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTT 1865
Db      13034 ATCACGAATGTTTTTAATGGCATCTAAAATGGTGAATCCTTTCCAGAAGGTTTTCAATTT 13093

Qy       1866 ACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGG 1925
Db      13094 GCTTTGCCCAGATCCAT-----CAGAGACATCACTATCCATGGCAGCTAGAGCCTTATGA 13148

Qy       1926 AATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGC 1985
Db      13149 AATGTATTTCTCAAATAATAAGACTTGAAGGTCGAAATTACCCCTTGATCCATGGGCTGC 13208

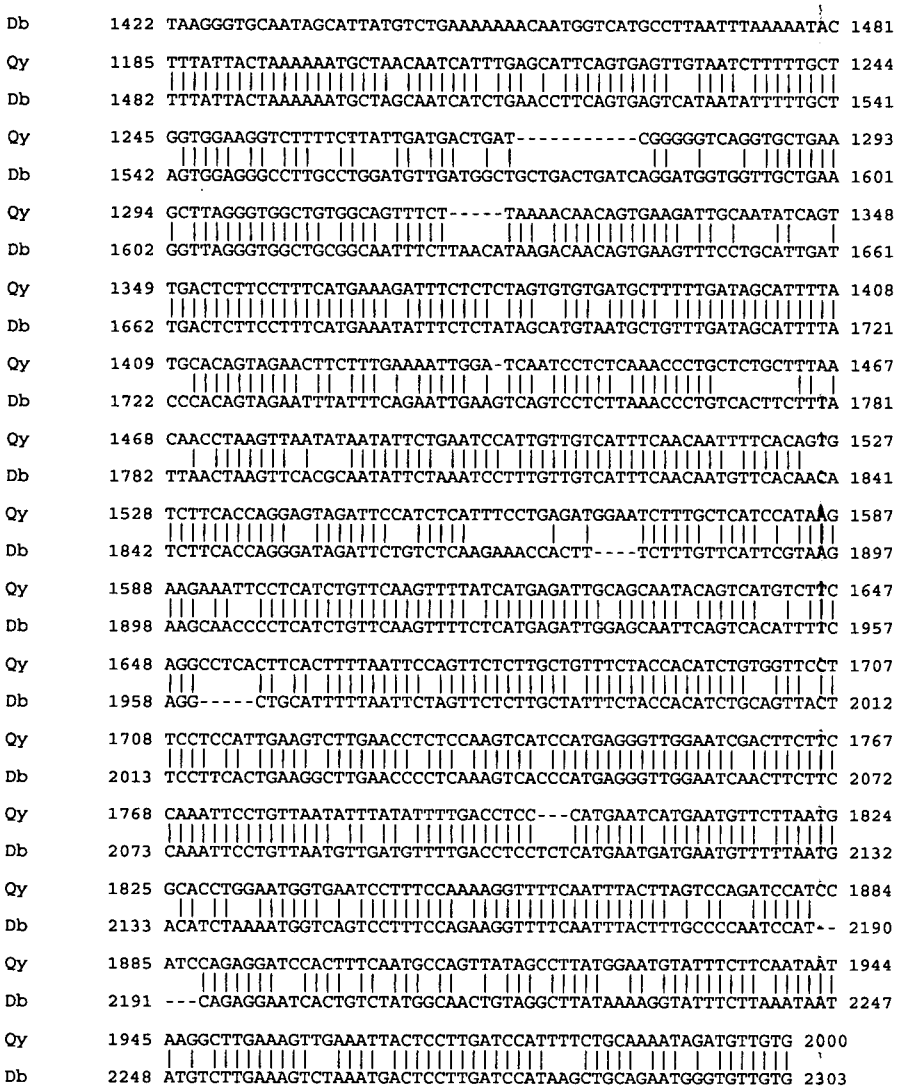
Qy       1986 AAAATAGATGTTGTG 2000
Db      13209 AGAATGGATGTTGTG 13223
```

RESULT 6
ADM01834
ID ADM01834 standard; cDNA; 3895 BP.
XX
AC ADM01834;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human cDNA of the invention SEQ ID NO:519.
XX
KW ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-723558/69.
DR P-PSDB; ADM04277.
XX
PT New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 519; 305pp; English.
XX
CC The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
CC cDNA sequence of the invention.
XX
SQ Sequence 3895 BP; 1139 A; 804 C; 807 G; 1145 T; 0 U; 0 Other;

Query Match 25.8%; Score 516; DB 11; Length 3895;
Best Local Similarity 77.5%; Pred. No. 1.8e-78;
Matches 741; Conservative 0; Mismatches 175; Indels 40; Gaps 8;

```
Qy       1071 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 1130
Db       1362 ATAAATTTTTTGGTTTCCCAGTACATGTAAAAGTTATAATTACACTATACTATAGTCTAT 1421

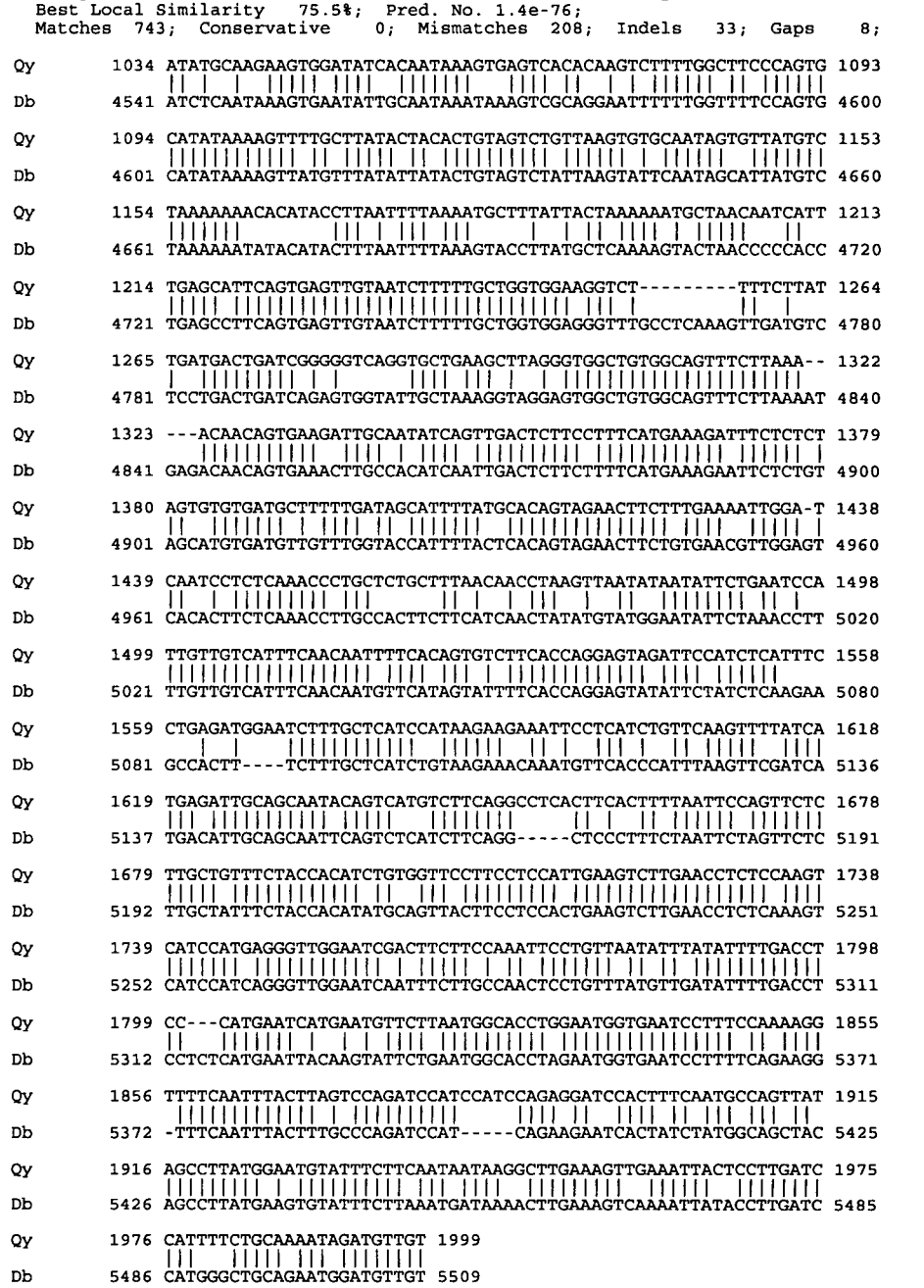
Qy       1131 TAAGTGTGCAATAGTGTTATGTCTAAAAAAA------CACATACCTTAATTTTAAAATGC 1184
```



RESULT 7
ABD32923_7
Continuation (8 of 8) of ABD32923 from base 700001 (Human cancer-associated genomic DNA
WP Sequence split into 8 fragments LOCUS ABD32923 Accession Abd32923

WP	Fragment Name	Begin	End
WP	ABD32923_0	1	110000
WP	ABD32923_1	100001	210000
WP	ABD32923_2	200001	310000
WP	ABD32923_3	300001	410000
WP	ABD32923_4	400001	510000
WP	ABD32923_5	500001	610000
WP	ABD32923_6	600001	710000
WP	ABD32923_7	700001	788759

Query Match 25.3%; Score 505.2; DB 13; Length 88759;
Best Local Similarity 75.5%; Pred. No. 1.4e-76;
Matches 743; Conservative 0; Mismatches 208; Indels 33; Gaps 8;



```
RESULT 8
ABD32923_6
Continuation (7 of 8) of ABD32923 from base 600001 (Human cancer-associated genomic DNA
WP Sequence split into 8 fragments  LOCUS ABD32923 Accession Abd32923
WP      Fragment Name      Begin     End
WP      ABD32923_0             1     110000
WP      ABD32923_1        100001     210000
WP      ABD32923_2        200001     310000
WP      ABD32923_3        300001     410000
WP      ABD32923_4        400001     510000
WP      ABD32923_5        500001     610000
WP      ABD32923_6        600001     710000
WP      ABD32923_7        700001     788759

  Query Match             25.3%;  Score 505.2;  DB 13;  Length 110000;
  Best Local Similarity   75.5%;  Pred. No. 1.4e-76;
  Matches  743;  Conservative    0;  Mismatches  208;  Indels   33;  Gaps    8;

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
             || |  |  ||||| ||||  |||||||  |||| ||  | | |||||| || ||||||
Db     104541 ATCTCAATAAAGTGAATATTGCAATAAATAAAGTCGCAGGAATTTTTTGGTTTTCCAGTG 104600

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
             |||||||||||| || ||||| || |||||||||| |||||| | ||||||  |||||||
Db     104601 CATATAAAAGTTATGTTTATATTATACTGTAGTCTATTAAGTATTCAATAGCATTATGTC 104660

Qy       1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
             |||||||         ||| | ||| |||      |  | || |||| | |||||   ||
Db     104661 TAAAAAATATACATACTTTAATTTTAAAGTACCTTATGCTCAAAAGTACTAACCCCCACC 104720

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT---------TTTCTTAT 1264
             ||||| ||||||||||||||||||||||||||||| ||| |           ||  |
Db     104721 TGAGCCTTCAGTGAGTTGTAATCTTTTTGCTGGTGGAGGGTTTGCCTCAAAGTTGATGTC 104780

Qy       1265 TGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-- 1322
             |  ||||||||| | |       |||| ||| |  | ||||||||||||||||||||||
Db     104781 TCCTGACTGATCAGAGTGGTATTGCTAAAGGTAGGAGTGGCTGTGGCAGTTTCTTAAAAT 104840

Qy       1323 ---ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
                 ||||||||||  |||| | |||| |||||||||| |||||||||||| |||||| |
Db     104841 GAGACAACAGTGAAACTTGCCACATCAATTGACTCTTCTTTTCATGAAAGAATTCTCTGT 104900

Qy       1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-T 1438
             ||  ||||||| | |||| || |||||||  |||||||||||||| ||||  ||||| |
Db     104901 AGCATGTGATGTTGTTTGGTACCATTTTACTCACAGTAGAACTTCTGTGAACGTTGGAGT 104960

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
             ||  | |||||||||| |||       || |  | |||   |  ||  |||||||| || |
Db     104961 CACACTTCTCAAACCTTGCCACTTCTTCATCAACTATATGTATGGAATATTCTAAACCTT 105020

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
             |||||||||||||||||| |||| ||| | |||||||||||||| |||| ||||||
Db     105021 TTGTTGTCATTTCAACAATGTTCATAGTATTTTCACCAGGAGTATATTCTATCTCAAGAA 105080

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
                |  |    ||||||||||| ||||||  || |  ||| |  || |||||  ||||
Db     105081 GCCACTT----TCTTTGCTCATCTGTAAGAAACAAATGTTCACCCATTTAAGTTCGATCA 105136

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
             ||| |||||||||| |||||   ||||||||     || |  || |||||| |||||||
Db     105137 TGACATTGCAGCAATTCAGTCTCATCTTCAGG-----CTCCCTTTCTAATTCTAGTTCTC 105191

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
             ||||| |||||||||||| ||   ||| |||||||| ||||||||||||||||||| ||||
Db     105192 TTGCTATTTCTACCACATATGCAGTTACTTCCTCCACTGAAGTCTTGAACCTCTCAAAGT 105251

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
             ||||||| |||||||||||| | ||||| | || ||||||| || || ||||||||||||
Db     105252 CATCCATCAGGGTTGGAATCAATTTCTTGCCAACTCCTGTTTATGTTGATATTTTGACCT 105311

Qy       1799 CC---CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855
             ||   ||||||| |  | | |||| |||||||||| ||||||||||||||||| || ||||
Db     105312 CCTCTCATGAATTACAAGTATTCTGAATGGCACCTAGAATGGTGAATCCTTTTCAGAAGG 105371

Qy       1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915
             |||||||||||||| | |||||||||||     |||| ||  |||| || ||| ||| ||
Db     105372 -TTTCAATTTACTTTGCCCAGATCCAT-----CAGAAGAATCACTATCTATGGCAGCTAC 105425

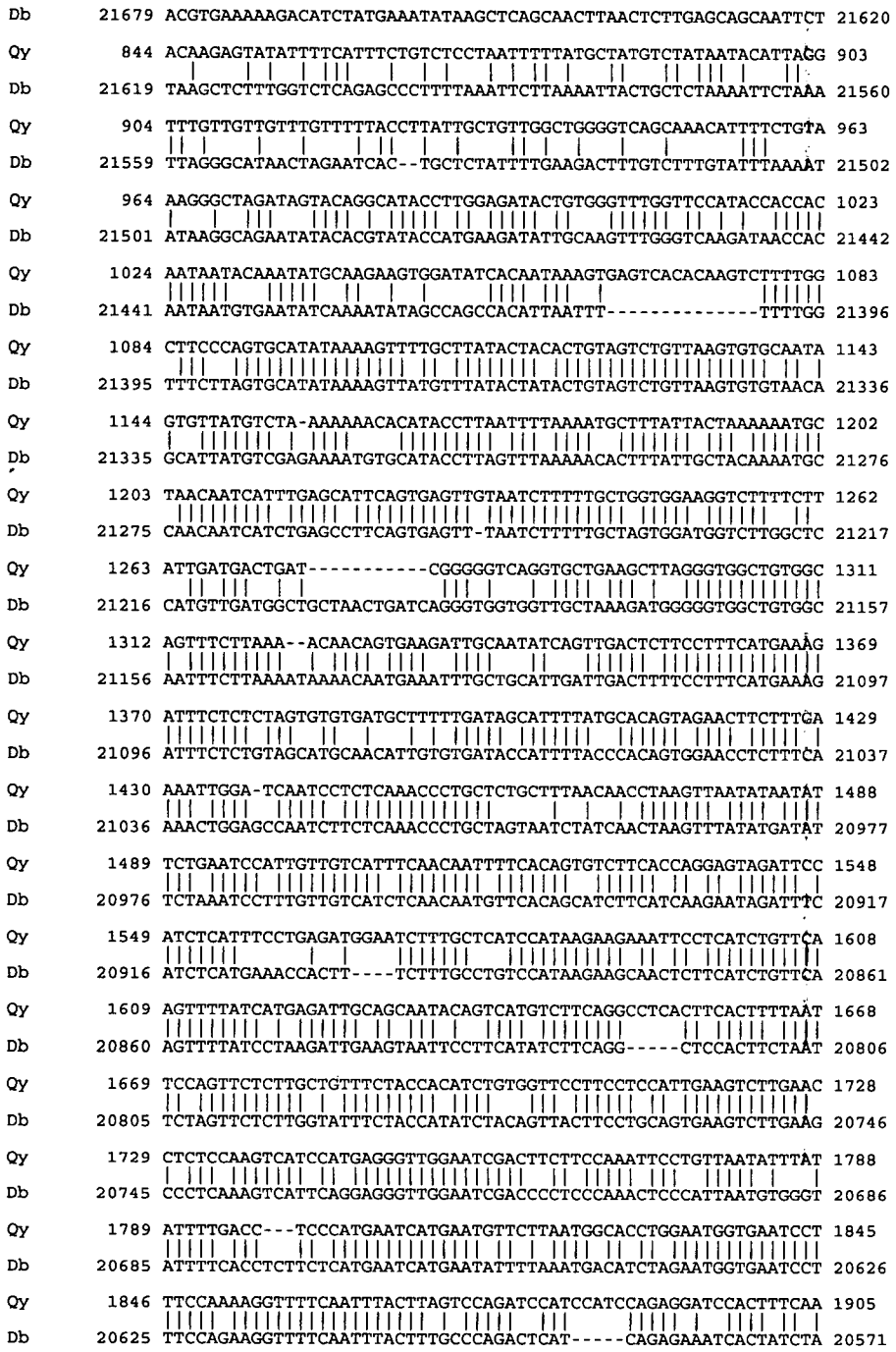
Qy       1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
             ||||||||| | |||||||||| ||| ||||  |||||||||  ||||||  ||||||||
Db     105426 AGCCTTATGAAGTGTATTTCTTAAATGATAAAACTTGAAAGTCAAAATTATACCTTGATC 105485

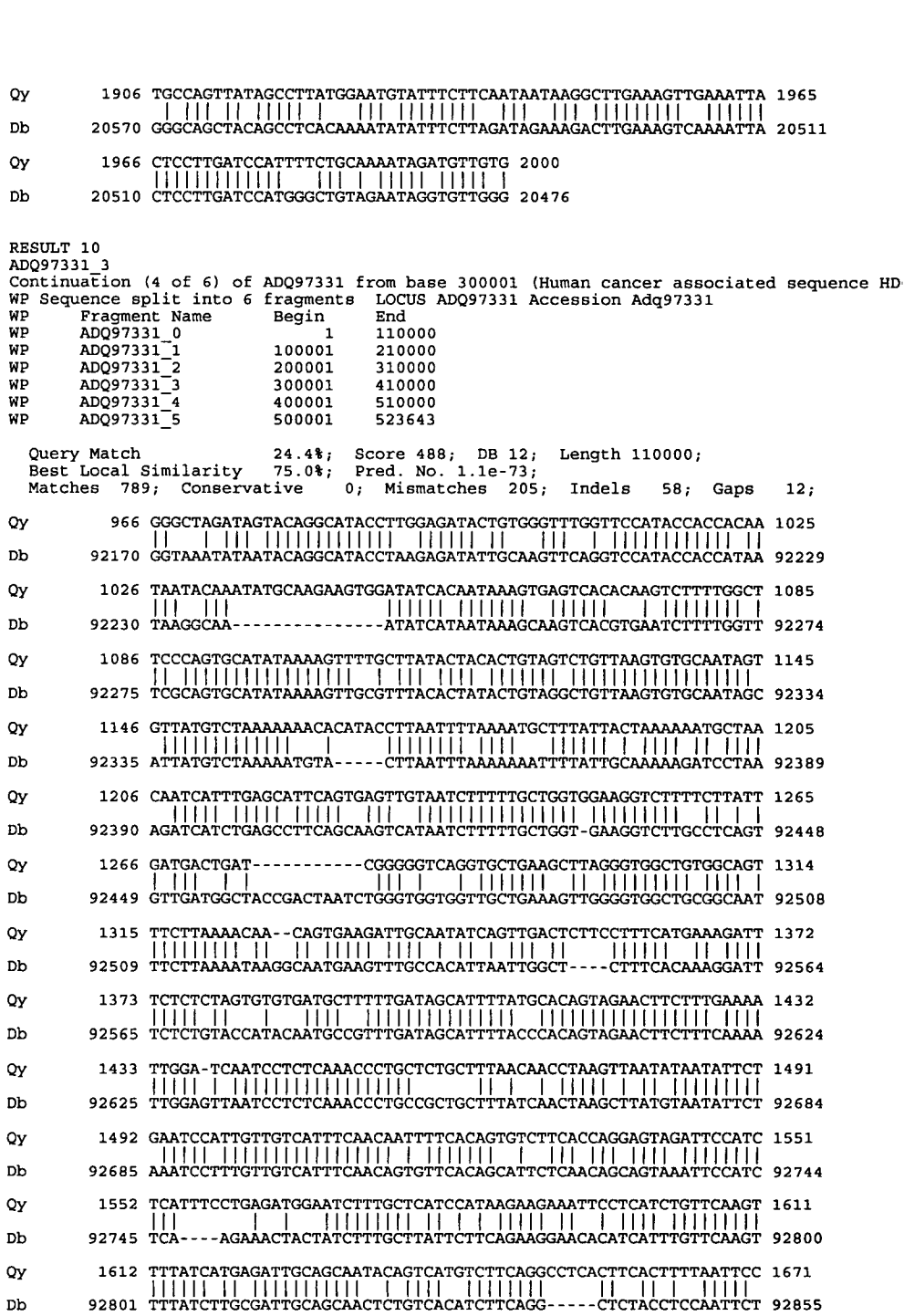
Qy       1976 CATTTTCTGCAAAATAGATGTTGT 1999
             |||   ||||| ||| ||||||||
Db     105486 CATGGGCTGCAGAATGGATGTTGT 105509


RESULT 9
ABD33512/c
ID   ABD33512 standard; DNA; 55829 BP.
XX
AC   ABD33512;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-100.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 688; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 55829 BP; 17201 A; 9903 C; 10825 G; 17900 T; 0 U; 0 Other;

  Query Match             25.0%;  Score 500.6;  DB 13;  Length 55829;
  Best Local Similarity   69.4%;  Pred. No. 8.1e-76;
  Matches  857;  Conservative    0;  Mismatches  329;  Indels   49;  Gaps   11;

Qy        784 ACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAACCTT 843
              || |   | |||   |  | |||||  |  | | || || || | |   | || ||   |
```





RESULT 10
ADQ97331_3
Continuation (4 of 6) of ADQ97331 from base 300001 (Human cancer associated sequence HD
WP Sequence split into 6 fragments LOCUS ADQ97331 Accession Adq97331

	Fragment Name	Begin	End
WP	ADQ97331_0	1	110000
WP	ADQ97331_1	100001	210000
WP	ADQ97331_2	200001	310000
WP	ADQ97331_3	300001	410000
WP	ADQ97331_4	400001	510000
WP	ADQ97331_5	500001	523643

Query Match 24.4%; Score 488; DB 12; Length 110000;
Best Local Similarity 75.0%; Pred. No. 1.1e-73;
Matches 789; Conservative 0; Mismatches 205; Indels 58; Gaps 12;

```
Qy       1672 AGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTC 1731
             |||||||||||| |||| ||||||||||    || ||||||||| |||||| ||||||| |
Db      92856 AGTTCTCTTGCTATTTCCACCACATCTGCTATTACTTCCTCCACTGAAGTGTTGAACCCC 92915

Qy       1732 TCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATT 1791
             ||  |||||||||| |||||||||||| |||||||||| |||||||||| || |||||
Db      92916 TC--AGTCATCCATAAGGGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGATATT 92973

Qy       1792 TTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTC 1848
              ||||   ||||| ||||||    |||||| ||||||||| || |||||||||| |||||
Db      92974 TTGATCTCCTCCTATGAATTTCAAATGTTGTTAATGGCATCTAGAATGGTGAAGCCTTTT 93033

Qy       1849 CAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGC 1908
              || |||||||||||  |||| ||||||||||||     |||||||  |||| || |||
Db      93034 CAGAAGGTTTTCAATGGACTTTGTCCAGATCCAT-----CAGAGGAATCACTATCTATGG 93088

Qy       1909 CAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTC 1968
              ||| ||||||| ||   |||||||||||| ||||| |   |||||||||  | ||||| |
Db      93089 CAGCTATAGCCGTACAAAATGTATTTCTTAAATAAGACAACTTGAAAGTCAAGATTACAC 93148

Qy       1969 CTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||||||||   |  || |||  ||||||||
Db      93149 CTTGATCCATGGGCCACAGAATGAATGTTGTG 93180
```

RESULT 11
ABA20544/c
ID ABA20544 standard; DNA; 21224 BP.
XX
AC ABA20544;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 12875.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.

PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
PS Disclosure; SEQ ID NO 12875; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease;
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 21224 BP; 6415 A; 4173 C; 4443 G; 6193 T; 0 U; 0 Other;

Query Match 24.0%; Score 480.2; DB 5; Length 21224;
Best Local Similarity 73.9%; Pred. No. 2.3e-72;
Matches 713; Conservative 0; Mismatches 223; Indels 29; Gaps 7;

```
Qy      1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
             |||    |||||||||||||  ||| || |||| | |||||| |||||| |||| ||||||
Db     14728 AAGCAAGTATCACAATAAACAGAGCCATACAAATTTTTTGGTTTCCCATTGCACATAAAA 14669

Qy      1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC 1162
             |||  ||||| |||| ||||||| ||  ||| |||| |||||  |||||||  |||||
Db     14668 GTTAGGCTTACACTATACTGTAGCCTACTAAATGTGAAATAGCATTATGTCTTTAAAAAT 14609

Qy      1163 A-----CATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAG 1217
             |    |||| |||||||| ||||| ||||||| ||||||| ||||||  || || ||||
Db     14608 AATGAGCATAGCTTAATTTAAAAATACTTTATTGCTAAAAATTGCTAATGATTATCTGAG 14549

Qy      1218 CATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCG 1277
               |||| |||||  ||       |     ||         ||       || |||||||||
Db     14548 TCTTCACTGAGTCTTAGAGAGGCT-----TGCCTCAATGTTGACGGCTGCTGACTGATCA 14494

Qy      1278 GGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACAACAGTGAAGATT 1337
             | |    |||      |   | ||||||||||||| | ||| || |  |||||||| ||
Db     14493 GAGTAGTGGTTGCTGAAGGTTAGGTGGCTGTGGCAATCTCTCAAGATGACAGTGAAGTTT 14434

Qy      1338 GCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTG 1397
               |   ||  | |||||||||||| |  | ||| ||  ||||||| |||||||||  |||
Db     14433 TCTGCATTGATGGACTCTTCCTTTAACAATAGAGTTTGCTCTAGTATGTGATGCTGCTTG 14374

Qy      1398 ATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAACCCTGC 1457
             | |||||||||  |||||  |||||||  |||  |||||    |||||||  ||||||||||
Db     14373 AGAGCATTTTACTCACAGCAGAACTTTTTTCAAAATCAAGTCAATCCTTTCAAACCCTGC 14314

Qy      1458 TCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAAT 1517
             |      ||  | ||  ||||||| | |||||||||  ||  || ||||||||||||||||||
Db     14313 TGCTACTTTATAAACTAAGTTTACATAATATTCTAAACCCTTTGTTGTCATTTCAACAAT 14254

Qy      1518 TTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCT 1577
             |||||||  ||||| |||| |||||||| |||||||        |  |    ||||||||
Db     14253 GTTCACAGCATCTTCGCCAGTAGTAGATTTCATCTCAAGAAACCATTT----TCTTTGCT 14198

Qy      1578 CATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAG 1637
             ||| || ||||| || || |||||  ||||| |||||||||| |||||||||||||| |||
Db     14197 GATCTATCAGAAGCAACTCTTCATCCATTCAAATTTTATCATTAGATTGCAGCAATTCAG 14138

Qy      1638 TCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATC 1697
             |||| |||||||||| ||||||     |||||| | |||||||||| ||||||||||||
Db     14137 TCATATCTTCAGGCCCCACTTC-----TAATTCTAATTCTCTTGCTATTTCTACCACATT 14083

Qy      1698 TGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAAT 1757
             ||  ||| | ||||||| |||||||||||||||||| || |||| ||| | ||||||||||
Db     14082 TGCAGTTACGTCCTCCACTGAAGTCTTGAACCTCTCAAACTCATTCATAAAGGTTGGAAT 14023

Qy      1758 CGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATCATGAAT 1814
             | |||||||||||| ||||| || | || |||||||||    |||||| ||| |||  |||
Db     14022 CAACTTCTTCCAAACTCCTGCTATTTTTGATATTTTGACCTCCTCCCTTGAGTCACAAAT 13963

Qy      1815 GTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCC 1874
             |||||||||||||  || ||||    |||  |||||| |||||||||| ||||||| ||||
Db     13962 GTTCTTAATGGCATCTAGAATACCAAATTGTTTCCAGAAGGTTTTCAGTTTACTTTGTCC 13903

Qy      1875 AGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTT--ATGGAATGTAT 1932
             ||||||||      ||||||   |||| || ||  |||  | |||  ||  |    ||||||
Db     13902 AGATCCAT-----CAGAGGGATCACTATCTATAGCAGCTCTAGGTTTGCAAAACATGTAT 13848

Qy      1933 TTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAG 1992
             |||||  | |||  |||  |||| |||| |||  |  ||||||    |||||    || || |||||
Db     13847 TTCTTAAGTAACAAGACTTGCAAGTCGAAGTAACTCCTCAATCCAGGGGCTCCAGAATAG 13788

Qy      1993 ATGTT 1997
             |||  |
Db     13787 ATGCT 13783
```

RESULT 12
AAL04361/c
ID AAL04361 standard; DNA; 32192 BP.
XX
AC AAL04361;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 7049.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.

PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 7049; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 32192 BP; 9426 A; 6338 C; 6511 G; 9917 T; 0 U; 0 Other;

Query Match 24.0%; Score 480; DB 4; Length 32192;
Best Local Similarity 73.9%; Pred. No. 2.5e-72;
Matches 776; Conservative 0; Mismatches 220; Indels 54; Gaps 11;

```
Qy       971 AGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATA 1030
             | | | ||||||||| ||||   ||||| |||||||||  || || |  |    |||||
Db      8208 ACAAAATACAGGCACACCTCAAAGATAATGTGGGTTTCATTTCAGATAATGGAAATAAAG 8149

Qy      1031 CAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCA 1090
               |||||           ||  | |  ||| || |||||  || |   |||| ||||||
Db      8148 TGAATAT----------TGTAATAAAGCAAATTGTGTCACGTAAATTGCTTGGTTTCCCA 8099

Qy      1091 GTGCATATAAAAGTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTAT 1150
             |||||||||||||||  |    || ||||||||||||||| |||| ||||||||||  ||||
Db      8098 GTGCATATAAAAGTTATATCTACACTACACTGTAGTCTATTAAATGTGCAATAGAATTAT 8039

Qy      1151 GTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC 1210
               ||||||| |      | ||||||  |   || ||||||||||||||| |||||||||
Db      8038 ATCTAAAATA----TGTACATAATTTAAGCATACTTTATTACTAAAAAA-GCTAACAATC 7984

Qy      1211 ATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGA 1270
             || |||||  || |||||||||||||||| |||||||||| || |||   | | || |||
Db      7983 ATCTGAGCCTTTAGTGAGTTGTAATCTTGTTGCTGGTGGAGGGGCTTGCTTCAATGTTGA 7924

Qy      1271 CTGAT----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT 1319
               | |            ||| |   | |||||||| || | ||||||||||| |||
Db      7923 TGGCTGCTGACTGATCAGGGTGGTGGCTGCTGAAGGTTGGTGTGGCTGTGGCAATTTTAA 7864

Qy      1320 AAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTC 1374
             |||     ||||||  |||| |||| | |||| ||||||| |  ||| ||||||  |||||
Db      7863 AAAAGAAGACAACAACGAAGTTTGCCACATCAATTGACTCCTTTTTTTATGAAAAATTTC 7804

Qy      1375 TCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATT 1434
             |||  ||  || |||||| ||||||||||||||  ||  |||||||||||| ||||||
Db      7803 TCTGAAGGCTGCGATGCTGTTTGATAGCATTTTACCCATGGTAGAACTTCTTTCAAAATT 7744

Qy      1435 GGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGA 1493
             ||| |||||||||| |||||||       || |  | ||||||| || |||||||||  |
Db      7743 GGAGTCAATCCTCTCCAACCCTGTCACTGCTTTATCAACTAAGTTTATGTAATATTCTAA 7684

Qy      1494 ATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTC 1553
             |||| || ||||||||||||||||| |||||||  |||||||||||| || |||||||||
Db      7683 ATCCCTTCTTGTCATTTCAACAATGTTCACAGCATCTTCACCAGGAGGAGTTTCCATCTC 7624

Qy      1554 ATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTT 1613
             |    | |  |    ||||||| |||||||||||| || ||| ||||| || ||||||
Db      7623 AAGAAATCACGT----TCTTTGCACATCCATAAGAAGCAACTCCCCATCTATTTAAGTTT 7568

Qy      1614 TATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAG 1673
             |||||||||  | |||||||| || |||  ||||||||      | ||||| | |||| |
Db      7567 TATCATGAGTTGGCAGCAATTCACTCACATCTTCAGG-----TTCCACTTATGATTCTAA 7513

Qy      1674 TTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTC 1733
             |||||||||| |||| |||| |||||   ||| |||||||||         |||| | |||
Db      7512 TTCTCTTGCTATTTCCACCATATCTGCAGTTACTTCCTCCA--------CTGAATCCCTC 7461

Qy      1734 CAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTT 1793
               ||||||  ||||||| |||||||| |||||||||||| |||||||||| || || | ||
Db      7460 AAAGTCACTCATGAGGATTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGATGTATT 7401

Qy      1794 GA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
             ||   ||| |||||||||| ||||||| || ||||||     ||| |||||  |||||||
Db      7400 GACATCCTTCCATGAATCACGAATGTTTTTCATGGCATTGACAATAGTGAACTCTTTCCA 7341

Qy      1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910
             |||||||||||||||||| | | |||||||||       |  |    ||| |  | | ||
Db      7340 GAAGGTTTTCAATTTACTTTGCCTAGATCCATCAGATAGGCAGGAGGACTGT--AAGGCA 7283

Qy      1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970
             | |||||| |||   |||||||||||| |||||||| ||||||||||||||||||||||
Db      7282 GGTATAGCTTTACAAAATGTATTTCTTAAATAATAAGACTTGAAAGTTGAAATTACTCCT 7223

Qy      1971 TGATCCATTTTCTGCAAAATAGATGTTGTG 2000
             |||||||   ||||| ||| ||  |||||
Db      7222 TGATCCATGGGCTGCAGAATGGACATTGTG 7193
```

RESULT 13
ACN44010/c
ID ACN44010 standard; DNA; 198522 BP.
XX
AC ACN44010;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG1643869.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 244; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable

```
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 198522 BP; 58663 A; 35063 C; 36453 G; 68323 T; 0 U; 20 Other;

  Query Match             24.0%;  Score 479.6;  DB 11;  Length 198522;
  Best Local Similarity   70.9%;  Pred. No. 3.1e-72;
  Matches  812;  Conservative    0;  Mismatches  279;  Indels   55;  Gaps   11;

Qy        877 TTTATGCTATGTCTATAATACATTAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTT 936
              | || | |     || |||| | |  || | |   || || | ||    |      |
Db     131071 TATAAGTTGAAAATAAAATAAAATCTGTATCTACTAAATTTTTCTCACTCATTCAGATTC 131012

Qy        937 GGCTGGGGTCAGCAAACATTTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGAT 996
                |  | |    |||  | |   |  | ||  |     ||||||||| |||  ||||
Db     131011 ACCAAAGATTTAAAAATGTCTAAGGATACAGGAAAACATATACAGGCATCCCTGAGAGAC 130952

Qy        997 ACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAAGAAGTGGATATCACA 1056
              | ||   |||||||||| |||||| ||||||                | | |||||| |
Db     130951 ATTGCAAGTTTGGTTCCAGACCACCCCAATAAG---------------GAGAATATCATA 130907

Qy       1057 ATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACT 1116
              | ||||| |||||||  | | |||||| ||||||   |||||||||| | || ||| |||
Db     130906 ACAAAGTAAGTCACATGAATTTTTTGGTTTCCCAACACATATAAAAGCTATGTTTACACT 130847

Qy       1117 ACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC-----ACATACCTT 1171
              | |||| ||||| ||||||| |||| ||  |||||||| ||||  |     | |||||||
Db     130846 ATACTGCAGTCTATTAAGTGGGCAACAGCATTATGTCTCAAAAGCCAATGTATATACCTT 130787

Qy       1172 AATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTG 1231
              || || |||||  |||||| ||||||| |  ||  ||||| ||||| ||||||||||
Db     130786 AACTTGAAAATATTTTATTGCTAAAAAAGGTGAATGATCATCTGAGCCTTCAGTGAGT-- 130729

Qy       1232 TAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA-------TGACTGATCGGGGGTCA 1284
              | ||| |||||||||| || ||||||  || | |||       |||||||| |||
Db     130728 CACTCTCTTTGCTGGTAGAGGGTCTTGCCTCAATGATGACTGCTGACTGATCAGGGTGGT 130669

Qy       1285 GGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGC 1339
              |||    ||  || | |||||||||||| |||||||||     || ||  ||||| ||||
Db     130668 GGTTGCTAACGTTGGAGTGGCTGTGGCAATTTCTTAAAATAAGACGACCATGAAGTTTGC 130609

Qy       1340 AATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGAT 1399
                |||||  |||||||||||||||  | |||||||||| | |  |  |||||| ||||||
Db     130608 TCTATCATGTGACTCTTCCTTTCACAATAGATTTCTCTGTGGCATAAGATGCTCTTTGAT 130549

Qy       1400 AGCATTTTATGCACAGTAGAACTTCTTTGAAAATT-GGATCAATCCTCTCAAACCCTGCT 1458
              |||||||||  ||   || |||||||||| |||||| || |||| |||||||||| ||||
Db     130548 AGCATTTTACCCAACATAAAACTTCTTTCAAAATTGGGGTCAAGCCTCTCAAACTCTGCC 130489

Qy       1459 CTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATT 1518
                || |    ||||||| || ||||||||| | ||| ||||||||||||||||||||
Db     130488 GCTGCTTCATGCACTAAGTTTATTTAATATTCTAACTCCTTTGTTGTCATTTCAACAATG 130429

Qy       1519 TTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTC 1578
              |||||||  |||| || | || ||| | | ||||||       |  |    |||||||||
Db     130428 TTCACAGCATCTTTACTAAGAATAGTTCCTATCTCAAGAAACCACTT----TCTTTGCTC 130373

Qy       1579 ATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGT 1638
              |||||||| | | || |||||||||||||||||| | ||||||||||||||||  |||
Db     130372 ATCCATAAAATGCAACTCCTCATCTGTTCAAGTGTCCTCATGAGATTGCAGCAATTTAGT 130313

Qy       1639 CATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCT 1698
              || ||||| ||       ||   ||  | |  |||||||||  | |||| |||||| ||
Db     130312 CACGTCTTAAG-------CTCCACTTCTAATTTAGTTCTCTTATTATTTCCACCACAGCT 130260

Qy       1699 GTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCC-ATGAGGGTTGGAAT 1757
                 ||| |||||||||| |||| |||| |  | |||  |||||||| ||||||| ||||||
Db     130259 ACAGTTACTTCCTCCACTGAACTCTTAACTCCCTCAGAGTCATCCAATGAGGGATGGAAT 130200

Qy       1758 CGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATCATGAAT 1814
                | |   |||||||| |||||| | | || ||||| |||   ||||||||||||||||||
Db     130199 CAATGACTTCCAAACTCCTGTAAGTGTTGATATTCTGACCTCCTCCCATGAATCATGAAT 130140

Qy       1815 GTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCC 1874
              |||||||||||||  || ||| |||||||||||||| |||||||| |||  |||  ||||
Db     130139 GTTCTTAATGGCATCTAGAACGGTGAATCCTTTCCAGAAGGTTTTTAATCGACTGTGTCC 130080

Qy       1875 AGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTT 1934
              |||| |||     ||||||| | || || ||  |||||||||||||| || |||||| ||
Db     130079 AGATGCAT-----CAGAGGACTCGCTATCTATAGCAGTTATAGCCTTGTGAAATGTACTT 130025

Qy       1935 CTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGAT 1994
              ||| | |||| || |||| |||| |||||||||||||| || ||||   |||| ||| |||
Db     130024 CTTAAGTAATGAGACTTGGAAGTAGAAATTACTCCTTTATTCATTGGATGCAGAATGGAT 129965

Qy       1995 GTTGTG 2000
              ||| ||
Db     129964 GTTTTG 129959
```

```
RESULT 14
ADA53284/c
ID   ADA53284 standard; cDNA; 2448 BP.
XX
AC   ADA53284;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human coding sequence, SEQ ID 852.
XX
KW   Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW   Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW   inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS   Homo sapiens.
XX
PN   EP1293569-A2.
XX
PD   19-MAR-2003.
XX
PF   21-MAR-2002; 2002EP-00006586.
XX
PR   14-SEP-2001; 2001JP-00328381.
PR   24-JAN-2002; 2002US-0350435P.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-395539/38.
DR   P-PSDB; ADA54923.
XX
PT   New polynucleotides encoding full-length polypeptides, e.g. secretory
PT   and/or membrane proteins, useful for developing medicines for diseases in
PT   which the gene is involved, or as target molecules for gene therapy.
XX
PS   Claim 1; SEQ ID NO 852; 205pp; English.
XX
CC   The present invention relates to novel human secretory or membrane
CC   proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC   ADA54071). The coding sequences are useful in the gene therapy of
CC   diseases caused by abnormalities of the proteins, e.g. cancer,
CC   inflammatory diseases, osteoporosis or neurological disease.
XX
```

```
SQ   Sequence 2448 BP; 734 A; 518 C; 542 G; 654 T; 0 U; 0 Other;

  Query Match             23.9%;  Score 477.6;  DB 10;  Length 2448;
  Best Local Similarity   79.5%;  Pred. No. 5.8e-72;
  Matches  676;  Conservative    0;  Mismatches  139;  Indels   35;  Gaps    8;

Qy       1172 AATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC-ATTTGAGCATTCAGTGAGTT 1230
             ||||| ||||| |||||||| |||||||||||| | ||| || ||||| ||| ||||||
Db       2447 AATTTAAAAATACTTTATTGCTAAAAAATGCTGATTATCAATCTGAGCCTTCGGTGAGTC 2388

Qy       1231 GTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-----------CGGG 1279
             ||| ||||||||||||||||| ||||||  ||    | |||  | |           |||
Db       2387 GTACTCTTTTTGCTGGTGGAGGGTCTTGCCTCGAGGTTGATGGTTGCTGACTGGTCAGGG 2328

Qy       1280 GGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAG 1334
              |   | |||||||| || |||||||||||||| |||| ||||     ||| | ||||||
Db       2327 TGGTGGCTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCCTAAAATAAGACATCGGTGAAG 2268

Qy       1335 ATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTT 1394
              ||     || |||||||||||||||||||||||||||||||| |||  || ||||||| |
Db       2267 CTTTTCGCATGAGTTGACTCTTCCTTTCATGAAAGATTTCTCTGTAGCATGCGATGCTGT 2208

Qy       1395 TTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACC 1453
              |||||||||||||   ||||||||| ||||||| ||||||||| ||||||||||||||||
Db       2207 TTGATAGCATTTTGCCCACAGTAGAGCTTCTTTCAAAATTGGAGTCAATCCTCTCAAACC 2148

Qy       1454 CTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAA 1513
              |||||     || |  | ||||||| || ||||||||| ||||| ||||||||||||||||
Db       2147 CTGCTGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGTTGTCATTTCAA 2088

Qy       1514 CAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTT 1573
              |||| |||||||  |||||||||||| |||||||||||||||        |  |    ||||
Db       2087 CAATGTTCACAGCATCTTCACCAGGTGTAGATTCCATCTCAAGAAACCACTT----TCTT 2032

Qy       1574 TGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAAT 1633
              ||| ||||||||||||  || ||||||||||| ||||||||||||||||||||||||||||
Db       2031 TGCGCATCCATAAGAAACAACTCCTCATCTGTTAAAGTTTTATCATGAGATTGCAGCAAT 1972

Qy       1634 ACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCA 1693
              ||||  |  ||||||||      ||  |||| |||||| |||||| ||||| |||| ||||
Db       1971 TCAGTTACATCTTCAGG-----CTCTACTTCTAATTCTAGTTCTTTTGCTATTTCCACCA 1917

Qy       1694 CATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTG 1753
              ||||||   || |||||||| |||||||||||||| |||  ||||||| ||||||||.|
Db       1916 CATCTGCACTTACTTCCTCCACTGAAGTCTTGAACCCCTCATAGTCATCAATGAGGGTGG 1857

Qy       1754 GAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATCAT 1810
              ||||| |||||||||||| |||||||||| || | | |||||   |||||||||||||
Db       1856 GAATCAACTTCTTCCAAACTCCTGTTAATGTTGACAATTTGACCTCCTCCCATGAATCAC 1797

Qy       1811 GAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTA 1870
              ||||||| ||||||||| ||  |||||||||||||||||| |||||||||||||| |||
Db       1796 GAATGTTTTTAATGGCATCTAAAATGGTGAATCCTTTCCAGAAGGTTTTCAATTTGCTTT 1737

Qy       1871 GTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGT 1930
              | ||||||||||     |||||    |||| || ||| ||| || |||||||||| |||||
Db       1736 GCCCAGATCCAT-----CAGAGACATCACTATCCATGGCAGCTAGAGCCTTATGAAATGT 1682

Qy       1931 ATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAAT 1990
              ||||||  |||||||| |||||| || |||||||| |||||||||||    ||||| |||
Db       1681 ATTTCTCAAATAATAAGACTTGAAGGTCGAAATTACCCCTTGATCCATGGGCTGCAGAAT 1622

Qy       1991 AGATGTTGTG 2000
              ||||||||||
Db       1621 GGATGTTGTG 1612


RESULT 15
ACN44350/c
ID   ACN44350 standard; DNA; 276276 BP.
XX
AC   ACN44350;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG17121.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 754; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 276276 BP; 68379 A; 69211 C; 66764 G; 71922 T; 0 U; 0 Other;

  Query Match             23.8%;  Score 475.8;  DB 11;  Length 276276;
  Best Local Similarity   74.5%;  Pred. No. 1.4e-71;
  Matches  729;  Conservative    0;  Mismatches  217;  Indels   33;  Gaps    9;

Qy       1040 AAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATA 1099
              ||  ||   ||||  |||| ||||  ||||||| || | |||||  || |||||| |||||
Db      40173 AAAGAGCAAATATTGCAATGAAGCAAGTCACATAAATTTTTTGATTTTCCAGTGAATATA 40114

Qy       1100 AAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCT-AAAA 1158
              |||||| || || ||  |     |||||| |||||| || |||||   ||||||| ||||
Db      40113 AAAGTTATGTTTGTATCA-----TAGTCTATTAAGTATGTAATAGCTCTATGTCTCAAAA 40059

Qy       1159 AAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGC 1218
              |   ||||| |||||||  |||||||||||||| | |||| |||||||| || || |||||
Db      40058 ATGTACATATCTTAATTAAAAAATGCTTTATTGCCAAAATATGCTAACGATTATCTGAGC 39999

Qy       1219 ATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGAT---------- 1268
              ||||| |||||||||||||||||||||||||  |||||  |||  || |
Db      39998 TTTCAGGGAGTTGTAATCTTTTTGCTGGTGGAGAGTCTTGCCTTGGTGTTGGATGGCTGC 39939

Qy       1269 --GACTGATCGGGGGTCAGG-TGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACA 1325
                |||||||| ||| |  || | |||||||||| ||||| |||||||| ||| ||||  ||
Db      39938 TAGACTGATCAGGGTTGTGGTTACTGAAGCTTGGGGTGACTGTGGCAATTTGTTAAGGCA 39879

Qy       1326 ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTG 1385
              ||| ||||  | || |  |   ||||||||| | || |   ||||||||| | ||   ||
Db      39878 ACAATGAAATTGGCTAGCTTGATTGACTCTTTCGTTAACACAAGATTTCTGTGTATCATG 39819

Qy       1386 TGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCC 1444
```

```
Db      39818 CATTGCTGTTTGATAGATTTTTACCCACAGTAGAACTTCTTAAAAAATAGGAGTCAATCC 39759

Qy       1445 TCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTG 1504

Db      39758 TCTCAAACCCAGTCACTGCTTTATCAACTCAGTTTATGGAATATTCTAAATCCTTTGTTG 39699

Qy       1505 TCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGA 1564

Db      39698 TCATTTCAACAATATTCACAGCATCTTCACCAAGAGTAGTTTCCATCTCAAGAAACCACT 39639

Qy       1565 TGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGAT 1624

Db      39638 T----TCTTTGCTCAT-CATAAGAAGCAACCCCTCATCCATTCAAGTTTGATCATGAGGT 39584

Qy       1625 TGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTG 1684

Db      39583 TGCAGCAATTCAGTCACATCTTCAGG-----CTCCATTTCTAATTCTAGTTCTCTTGCTA 39529

Qy       1685 TTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCA 1744

Db      39528 TTTCCCCCGCATCTGCAGTTCCTTCCTCCACTGAAGCCTTGAACCCCTCAAAGTCATCCA 39469

Qy       1745 TGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCC 1801

Db      39468 TGAGGGTTGGAATCCATTTCTTCCAAACTCCTGTTCATGTTGCTATTTCAACCCCCTCCC 39409

Qy       1802 ATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCA 1861

Db      39408 ATGAATCACAAATGTTCTTCATGGAATCTAAAATGGTGACTCCTTTCCAGAAGGAAAGGA 39349

Qy       1862 ATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTT 1921

Db      39348 TTTCTATTTAATTTGCCTAGATCTATCAGAGGAATCACCATCTATGGCAGCTATAGCCTT 39289

Qy       1922 ATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTT 1981

Db      39288 ACAAAATGTATTTCTTAAAAAATAAGACTTGAGTGTCAGAATTACTCCCTGATCCATGTG 39229

Qy       1982 CTGCAAAATAGATGTTGTG 2000

Db      39228 CTGTAGAATGCAGGTTGTG 39210
```

RESULT 16
AAD46127
ID AAD46127 standard; DNA; 163350 BP.
XX
AC AAD46127;
XX
DT 27-DEC-2002 (first entry)
XX
DE Human tumour suppressor gene.
XX
KW Human; tumour suppressor protein; cell proliferative disorder; vaccine;
KW inflammation; brain cancer; adenocarcinoma; cervix cancer; bone cancer;
KW apoptosis; leukaemia; lymphoma; melanoma; therapy; chromosome 13; gene;
KW ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 2001. .160511
FT /*tag= a
FT /product= "Human tumour suppressor protein"
FT exon 2001. .2178
FT /*tag= b
FT intron 2179. .159807
FT /*tag= c
FT exon 159808. .160511
FT /*tag= d
XX
PN WO200268468-A2.
XX
PD 06-SEP-2002.
XX
PF 05-FEB-2002; 2002WO-US003235.
XX
PR 27-FEB-2001; 2001US-00793706.
XX
PA (PEKE) PE CORP NY.
XX
PI Gong F, Yan C;
XX
DR WPI; 2002-698658/75.
DR P-PSDB; AAE28633.
XX
PT New human tumor suppressor proteins, useful for developing human
PT therapeutic agents, or preventing or treating inflammation, or disorders
PT associated with cell proliferation, e.g. bone cancer, brain cancer,
PT leukemia or lymphoma.
XX
PS Claim 4; Fig 3; 200pp; English.
XX
CC The invention relates to human tumour suppressor polypeptides and
CC polynucleotides. Sequences of the invention are useful for diagnosing,
CC preventing or treating inflammation, or disorders associated with cell
CC proliferation and apoptosis e.g. bone cancer, brain cancer, cervix
CC cancer, adenocarcinoma, leukaemia, lymphoma or melanoma. They are
CC particularly useful as models for developing human therapeutic targets,
CC identifying therapeutic proteins, or serving as targets for the
CC development of human therapeutic agents that modulate tumour suppressor
CC protein activity in cells and tissues that express the tumour suppressor
CC protein. Polypeptides of the invention are used for identifying agents
CC that modulate their activity. They are useful for raising antibodies or
CC eliciting an immune response; as a reagent in assays designed to
CC quantitatively determine levels of the protein (or its binding partner or
CC ligand) in biological fluids; or as markers for tissues in which the
CC corresponding protein is preferentially expressed. The invention is also
CC used as vaccines. The present sequence is human tumour suppressor gene
CC located on chromosome 13
XX
SQ Sequence 163350 BP; 40430 A; 36356 C; 37002 G; 39020 T; 0 U; 10542 Other;

Query Match 23.5%; Score 469.4; DB 6; Length 163350;
Best Local Similarity 75.0%; Pred. No. 1.7e-70;
Matches 779; Conservative 0; Mismatches 201; Indels 59; Gaps 13;

```
Qy        978 ACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATAT 1037

Db     100173 ACAGGCAGACCTCAGACATACTGAAGGGTTGGTTCCAGACCACCACAATA---------- 100222

Qy       1038 GCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATA 1097

Db     100223 -----AAGCGAATATCACAATAAAGCAAGGCACACATATTTTCTGGTTTCCCAGTGCATA 100277

Qy       1098 TAAAAGTTTTGCTTA---TACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCT 1154

Db     100278 TAAAAGTTATGTTTACACTATTATACTGCAGTCTATTAAGTATATAATAGCATTATGTCT 100337

Qy       1155 AAAAAAAC-----ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAAT 1209

Db     100338 GAAAGAACAATGTACGTACCTTAATGAAAAAATATTTTATTGCTAACAAA-AATGATAAT 100396

Qy       1210 CATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTT---------TTC 1260

Db     100397 CATATGAGCCTTCAGTGA---GTCATCTTTTTGCTGGTGGAAGGTTTTGATTCAATATTA 100453

Qy       1261 TTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320

Db     100454 TGGCTGCTGACTGATCAAGGTAGTGGTTGATGAAGGCTGAGTTACTGTGGCAATTTCTTA 100513

Qy       1321 AA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCT 1375

Db     100514 AAATAAGACAATAATGAAGTCTGTGCCATCAATTGACTCTTCCTTTCATGAAAGATTTCT 100573
```

```
Qy       1376 CTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTG 1435
             || |||  |   | |  ||||||||||||||||  ||| | ||||  ||||| |||||||
Db     100574 CTGTAGCATCCAACG---TTTGATAGCATTTTACCCACGGCAGAATGTCTTTCAAAATTG 100630

Qy       1436 GA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAA 1494
              ||  || ||||||||||||||||      || |  | |||| || ||  |||| ||| ||
Db     100631 GAGCCAGTCCTCTCAAACCCTGCCACTGCTTTATCAACTAAATTTATGCAATAGTCTAAA 100690

Qy       1495 TCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCA 1554
              |||  |||||||||||||||||| | |||||  ||||||||||||||||||||||||||
Db     100691 TCCTCTGTTGTCATTTCAACAATGTCCACAGCATCTTCACCAGGAGTAGATTCCATCTCA 100750

Qy       1555 TTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTT 1614
                |  |    |||||||||||||||||||||| || ||||| | ||||||| ||||
Db     100751 AGAAACCACTT----TCTTTGCTCATCCATAAGAAGCAACTCCTCGTTTGTTCAACTTTT 100806

Qy       1615 ATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGT 1674
              |||||||| |||||||| | ||||||  ||| ||||   || |||   ||  | |  |||
Db     100807 ATCATGAGGTTGCAGCAGTTCAGTCACATCTCCAGGATCCATTTCCAATTCCAAT--AGT 100864

Qy       1675 TCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCC 1734
              ||||||||| || | |||| || ||  ||  ||||||||| ||| ||| |||||| |||
Db     100865 TCTCTTGCTATTCCCACCATATATGCAGTGACTTCCTCCAATGATGTCCTGAACCCCTCA 100924

Qy       1735 AAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTG 1794
              |||||| ||||||| |||||| |  ||||||| |||| |||||||| || |||||||||
Db     100925 AAGTCAACCATGAGAGTTGGACTTAACTTCTTTCAAACTCCTGTTAATGTTAATATTTTG 100984

Qy       1795 ACC---TCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAA 1851
              |||   |||||| ||||||  ||||||||| |||||  || ||||||||||||||||||
Db     100985 ACCTATTCCCATCAATCATACATGTTCTTACTGGCATCTAGAATGGTGAATCCTTTCCAG 101044

Qy       1852 AAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAG 1911
              ||||||| |||||||||| | |||| |||||      |||||||  |||| || ||| |||
Db     101045 AAGGTTTCCAATTTACTTTGCCCAGCTCCAT-----CAGAGGAATCACTATCTATGGCAG 101099

Qy       1912 TTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTT 1971
              |||||||||||| ||| | |||||| |||||||||| ||| ||||||||||||| ||||
Db     101100 CTATAGCCTTATGAAATATTTTTCTTAAATAATAAGACTTAAAAGTTGAAATTATTCCTT 101159

Qy       1972 GATCCATTTTCTGCAAAAT 1990
              |||| ||   ||||| |||
Db     101160 GATCAATGGGCTGCAGAAT 101178


RESULT 17
AAL52246_1
Continuation (2 of 4) of AAL52246 from base 100001 (Human genomic DNA for the gene encod
WP Sequence split into 4 fragments  LOCUS AAL52246 Accession Aal52246
WP      Fragment Name      Begin     End
WP      AAL52246_0             1    110000
WP      AAL52246_1        100001    210000
WP      AAL52246_2        200001    310000
WP      AAL52246_3        300001    378361

  Query Match           23.3%;  Score 466.6;  DB 10;  Length 110000;
  Best Local Similarity   72.7%;  Pred. No. 4.9e-70;
  Matches  758;  Conservative    0;  Mismatches  234;  Indels   51;  Gaps   10;

Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
              |||||||||||||||||||||| ||  |||||| ||||| |||||  ||||||  | ||||
Db      85373 TACAGGCATACCTTGGAGATATTGCTGGTTTGTTTCCAGACCACTGCAATAAAGCTAATA 85432

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096
              |                 ||||||||  ||||| | |      ||| |  ||||||||||
Db      85433 TTG---------------CAATAAAGCAAGTCATATAGACTGTTTTGTCTCCCAGTGCAT 85477

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156
              |||||||| |  ||| |||  |||||||||   ||||||||||||    |||||||||
Db      85478 ATAAAAGTTATATTTACACTCTACTGTAGTC---TAAGTGTGCAATGTCATTATGTCTTT 85534

Qy       1157 AAAAACACATACCT-TAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTG 1215
              |||||   ||   | | | ||||||    |||      |||||||||||||||||  |  |

Db      85535 AAAAATGTATGTATCTTAATTTAAACATATTTTACTGCAAAAAATGCTAACAATCGTCCG 85594

Qy       1216 AGCATTCAGTGAG---TTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACT 1272
              |||  ||||| |||   |  |||| ||||||||| |||| ||||||  || | || |||
Db      85595 AGCCTTCAGAGAGTCATAATAATATTTTTGCTGATGGAGGGTCTTGCCTCAATGTTGATG 85654

Qy       1273 GAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAA 1321
              | |              || |  || |||||||| ||||| |||||||||  | ||||||||
Db      85655 GCTGCTGACTGATCAAGGAGATAGTTGCTGAAGGTTAGGATGGCTGTGGAAATTTCTTAA 85714

Qy       1322 AACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAG 1381
              || || |          | |||  |||||||||||||||||| |||||||||| |||
Db      85715 AATAAGATAACGATGAAATTTGTCAACTGACTCTTCCTTTCATGAGAGATTTCTCTGTAG 85774

Qy       1382 TGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTG-GATCA 1440
              ||  ||||| |||||||||| ||||  |||||||||||| ||| |||||||  ||||
Db      85775 CATGCAATGCTATTTGATAGCATCTTATCTACAGTAGAACTTTTTTCAAAATTGTAATCA 85834

Qy       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500
              | |||| |||||||||      || |  | |||| ||  | | |||||||  ||| | |
Db      85835 GTTCTCTTAAACCCTGCTGCTGCTTTATCAACTAAATTTTTGTGATATTCTAAATTCTTG 85894

Qy       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560
              |||||||||||||||||  ||| ||  |||| |||||||||||||| |||||||
Db      85895 GTTGTCATTTCAACAATGCTCATAGCATCTTTACCAGGAGTAGATTTCATCTCA----AG 85950

Qy       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
              | |    |||||||| || ||||||||  |||| |||||||| |||| |||| ||||
Db      85951 AAACTACTTTCTTTGCTTATTCATAAGAAACAATTTCTCATCTGCTCAAATTTTAACATG 86010

Qy       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680
              |||||||||||||| ||||||  ||||||||     ||||| || |||||| ||| |||||
Db      86011 AGATTGCAGCAATTCAGTCACATCTTCAGG-----CTTCATTTCTAATTCTAGTACTCTT 86065

Qy       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740
              || |||| || |||||||  | | || |||| | ||||||||||||||| ||| ||||||
Db      86066 ACTATTTCCACTACATCTGCAGCTACTCCCTCTACTGAAGTCTTGAACCACTCAAAGTCA 86125

Qy       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACC--- 1797
              || ||||||| ||||||| |||| ||||||| || ||||| | |   ||| ||||||
Db      86126 TCTATGAGGGCTGGAATCAACTTATTCCAAACTCTTGTTAGTGTAGTTATCTTGACCTCT 86185

Qy       1798 TCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTT 1857
              ||| ||||||||  ||| |||||| || || || ||||||||||||| |  ||| ||||||
Db      86186 TCCAATGAATCACAAATTTTCTTACTGTCATCTAGAATGGTGAATCTTCCCCAGAAGGTT 86245

Qy       1858 TTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAG 1917
              |||||  |||||| | |||||| |     |  |||| |  |||| || ||| ||| |||||
Db      86246 TTCAACTTACTTTGCCCAGATTC-----TTGAGAGAAATCACTATCTATGGCAGCTATAG 86300

Qy       1918 CCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCA 1977
              |||| | ||| |||||||| | ||||||| || ||||||  |||| ||||||||||| ||
Db      86301 TCTTAAGAAATATATTTCTTAAGTAATAAGACTAGAAAGTCTAAATGACTCCTTGATTCA 86360

Qy       1978 TTTTCTGCAAAATAGATGTTGTG 2000
              |   ||| | ||| ||||| |||
Db      86361 TGAGCTGTAGAATGGATGTAGTG 86383


RESULT 18
ADQ59207
ID   ADQ59207 standard; DNA; 262090 BP.
XX
AC   ADQ59207;
XX
DT   09-SEP-2004  (first entry)
XX
DE   MSI-H carcinoma genomic DNA sequence SEQ ID NO:44.
XX
KW   coding mononucleotide repeat; cMNR; gene; antibody; MSI-H tumour;
KW   MSI-H carcinoma; high microsatellite instability tumour;
KW   high microsatellite instability carcinoma; cytostatic; ds.
XX
```

```
OS   Homo sapiens.
XX
PN   KR2004008012-A.
XX
PD   28-JAN-2004.
XX
PF   15-JUL-2002; 2002KR-00041304.
XX
PR   15-JUL-2002; 2002KR-00041304.
XX
PA   (KIMH/) KIM H G.
PA   (KIMN/) KIM N G.
PA   (LEEJ/) LEE J S.
PA   (RHEE/) RHEE H S.
XX
PI   Kim HG,  Kim NG,  Lee JS,  Rhee HS;
XX
DR   WPI; 2004-386326/36.
XX
PT   Genes containing coding mononucleotide repeats are useful in developing
PT   an antibody against MSI-H (hugh (sic high) microsatellite instability)
PT   tumor.
XX
PS   Claim 3; SEQ ID NO 44; 578pp; Korean.
XX
CC   The present invention describes genes containing coding mononucleotide
CC   repeats (cMNRs). The genes are useful for the development of an antibody
CC   against MSI-H (hugh microsatellite instability) tumour. Also described:
CC   (1) cDNA genes containing cMNRs with 10 or more nucleotide sequences, and
CC   selected from the cDNA genes having the nucleotide sequences of SEQ ID
CC   NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35,
CC   37, 39, 41 and 43; (2) cDNA genes, which are frameshift mutated by
CC   deletion or insertion of one or more base in the cMNRs; (3) genomic DNA
CC   genes containing cMNRs with 10 or more nucleotide sequences, and selected
CC   from the genomic DNA genes having the nucleotide sequences of SEQ ID
CC   NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36,
CC   38, 40, 42 and 44; and (4) genomic DNA genes, which are frameshift
CC   mutated by deletion or insertion of one or more base in the cMNRs. The
CC   genes have cytostatic activity. The present sequence represents an MSI-H
CC   carcinoma genomic DNA sequence from the present invention.
XX
SQ   Sequence 262090 BP; 74456 A; 48367 C; 51411 G; 86847 T; 0 U; 1009 Other;

  Query Match             23.3%;  Score 465.8;  DB 12;  Length 262090;
  Best Local Similarity   73.1%;  Pred. No. 6.9e-70;
  Matches  779;  Conservative    0;  Mismatches  232;  Indels   54;  Gaps   12;

Qy        972 GATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATAC 1031
              | || |||||| |||||||||||||||||| || |||||||||  | ||| ||||||  |
Db     118291 GCTAATACAGGAATACCTTGGAGATACTGAGGATTTGGTTCCAGGCTACCCCAATAAAGC 118350

Qy       1032 AAATATGCAAGAAGTGG----------ATATCACAATAAAGTGAGTCACACAAGTCTTTT 1081
              ||||||   | | |           ||||  |||||||| ||||||||  |   ||||
Db     118351 AAATATTGCAATAAAGCAGTAAATCAAATATTTCAATAAAGCGAGTCACATGAACTTTTT 118410

Qy       1082 GGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAA 1141
              || |||||||||| |||||||||| || ||| ||||  |||||||| |||||||| ||
Db     118411 GGTTTCCCAGTGCGTATAAAAGTTATGTTTACACTATGCTGTAGTCTATTAAGTGTGAAA 118470

Qy       1142 TAGTGTTATGTCTAAAAAAA---CACATACCTTAATTTTAAAATGCTTTATTACTAAAAA 1198
              |||  |||||||||||| |    |||||||||||||  || || ||| ||| |||||||
Db     118471 TAGCATTATGTCTAAAAACAATGTACATACCTTAATTAAAACATACTTCATTGCTAAAAA 118530

Qy       1199 ATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTT 1258
              |||||||| ||||| ||| | |||||   ||    | |||||||||||||||||| ||||||
Db     118531 ATGCTAACGATCATCTGAACCTTCAGCAAGCCAT-ATCTTTTTGCTGGTGGAGGGTCTTG 118589

Qy       1259 TCTTATTGATGACTGATCGGGGGTCA-----------GGTGCTGAAGCTTAGGGTGGCTG 1307
              ||   ||     | |   |  | |           | | ||| || || | |||  ||
Db     118590 CCTCCATGTCAGTGGCTGCTGACTGATCAGATTGGTGGTTACTGCAGGTTGGAGTGATTG 118649

Qy       1308 TGGCAGTTTCTTAAAACAACA----GTGAAGATTGCAATATCAGTTGACTCTTCCTTTCA 1363
              || || |||||||||| || |     ||||| |||  ||||||||||||||||  ||
Db     118650 TGACAATTTCTTAAAATAAGACAACATGAAGTTTGACATATCAGTTGACTCTTC--TTTC 118707

Qy       1364 TGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTT 1423
                 ||   |||||| ||   || |||||| |||||||||||||||   || ||||||||||
Db     118708 GTGAATGATTCTCTGTATCATGCGATGCTGTTTGATAGCATTTTACCCAGAGTAGAACTT 118767

Qy       1424 CTTTGAAAATTGGA-----TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTT 1478
              |||| |||||||||     ||  ||| ||||||||||||     ||| |  | |||||||
Db     118768 CTTTCAAAATTGGAGTCCGTCCTTCCCCTCAAACCCTGCCACTGTTTTATCAACTAAGTT 118827

Qy       1479 AATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGG 1538
                |||  ||    | |||   |||||||||||||||||| | || | ||  ||||||||||
Db     118828 --TATGATGGCGTAAATTTTTTGTTGTCATTTCAACAGTGTTTATAGCTTCTTCACCAGG 118885

Qy       1539 AGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCT 1598
              |||||||||  ||||||          | |    |  ||| | | |  ||| ||| || ||||
Db     118886 AGTAGATTCTATCTCAAAAAACCACTT----TTCTTGATTAACTGTAAAAAGTAACTCCT 118941

Qy       1599 CATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTT 1658
              ||||   |||||||| ||||||||||  |||||| |||||  ||||||||     ||
Db     118942 CATCCAGTCAAGTTTGATCATGAGATCACAGCAATTCAGTCACATCTTCAGG-----CTC 118996

Qy       1659 CACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGA 1718
              | ||| |||||| | |||||||||| ||   ||||  ||||  ||| |||||||||| |||
Db     118997 CTCTTATAATTCTAATTCTCTTGCTATTAAAACCATGTCTGCAGTTACTTCCTCCACTGA 119056

Qy       1719 AGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGT 1778
              |||||  ||||| ||| |||||||||||||||||||||||||  ||||||||||| ||||||
Db     119057 AGTCTCGAACCCCTCAAAGTCATCCATGAGGGTTGGAATCAGCTTCTTCCAAAGTCCTGT 119116

Qy       1779 TAATATTTATATTTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAAT 1835
              |||| || ||||||| |    |||||||||||||||||||||| |||||||     ||||
Db     119117 TAATGTTGATATTTTTATCTACTCCCATGAATCATGAATGTTCTGAATGGCATTCAGAAT 119176

Qy       1836 GGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGAT 1895
              || |||||||||||| || |||||| ||| | || | ||||    ||||| |   || ||
Db     119177 GGCGAATCCTTTCCAGAATGTTTTCCATTGATTTTGCCCAG----ATCCACCAGAAGAAT 119232

Qy       1896 CCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAA 1955
              | | | ||   || ||  | ||| |||||| |||||||||||| ||||||||  |||||||
Db     119233 CAATTATCTGTGGCAACTCTAGGCTTATGAAATGTATTTCTTAAATAATAAAACTTGAAA 119292

Qy       1956 GTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||||||  ||||||  ||||||   || || ||| |||  ||||
Db     119293 GTTGAAATGCCTCCTTCATCCATGGGCTACAGAATGGATAATGTG 119337


RESULT 19
ACN37213
ID   ACN37213 standard; DNA; 23271 BP.
XX
AC   ACN37213;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human periodontal disease related gene KRTHA2 SEQ ID NO:123.
XX
KW   periodontal disease; polymorphism; ds; human; gene; SNP;
KW   single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1796
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    1867
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    1964
FT                   /*tag=  c
```

```
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    4942
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    5020
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    5323
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,T"
FT   misc_feature    5841
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,T"
FT   misc_feature    6113
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    6132
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    6164
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,G"
FT   misc_feature    6347
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    6380
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,T"
FT   misc_feature    7316
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,G"
FT   misc_feature    7905
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    8025
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,A"
FT   misc_feature    8095
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    8252
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    8405
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    8838
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,A"
FT   misc_feature    8940
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    8985
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    9606
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    9763
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    9964
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    9979
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    9980
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    10001
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    10072
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    10105
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    10326
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide deletion,C"
FT   misc_feature    10920
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    11619
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    11796
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,C"
FT   misc_feature    12217
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    12218
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    12672
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    12954
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,G"
FT   misc_feature    13271
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    14010
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
```

```
FT   misc_feature    14011
FT                   /*tag=  an
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    14508
FT                   /*tag=  ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,C"
FT   misc_feature    14594
FT                   /*tag=  ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    14771
FT                   /*tag=  aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,T"
FT   misc_feature    14772
FT                   /*tag=  ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    15043
FT                   /*tag=  as
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    15126
FT                   /*tag=  at
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    15212
FT                   /*tag=  au
```

Query Match 23.3%; Score 465.6; DB 13; Length 23271;
Best Local Similarity 73.1%; Pred. No. 6.9e-70;
Matches 768; Conservative 0; Mismatches 224; Indels 58; Gaps 11;

```
Qy        972 GATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATAC 1031
              ||   |||||||||||||  |||||| ||   |||  |||| | ||||||||||||  |
Db      17926 GAAGATACAGGCATACCTCAGAGATAGTGCATGTTCAGTTCTAGACCACCACAATAAAGC 17985

Qy       1032 AAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAG 1091
              |||| |                 || ||||||||||||||||||   |||||| ||  | |
Db      17986 AAATGTA---------------ACCATAAAGTGAGTCACACAAACTTTTTGGTTTTTCTG 18030

Qy       1092 TGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATG 1151
              || || ||||  || |  ||| |||| ||| |||||| |||||||| | ||||  |||||
Db      18031 TGTATGTAAACATTATATTTACACTATACTATAGTCTATTAAGTGTACCATAGCATTATG 18090

Qy       1152 TCTAAAAAA----ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACA 1207
               |  |||||     | ||| ||||||||||||    ||  |    ||||||||||
Db      18091 GCCTAAAAACAATGTATATAACTTAATTTTAAAAATGCTTTATTGCTAAAAATGCTAAGG 18150

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              ||||| ||||| |||||  |||| |||||||||||||||||||||||||  || | ||
Db      18151 ATCATCTGAGCCTTCAGCAAGTTATAATCTTTTTGCTGGTGGAAGGTCTTGCCTCAATGT 18210

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              |||  | |            ||| |     ||||| |  |||| |||||||||||| |||
Db      18211 TGATGGCTGAGGACTGATCAGGGTGATGATTGCTGCAAGTTAGTGTGGCTGTGGCAATTT 18270

Qy       1317 CTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              || |||     ||| |||||||| ||||   ||||||||||||||||||||||||||||
Db      18271 CTGAAAATGAGACATCAGTGAAGTTTGCCGCATCAGTTGACTCTTCCTTTCATGAAAGAT 18330

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
              ||| || |||  |||||||| |||||||||||||||   |||||||| |||||| |||
Db      18331 TTCCCTGTAGCATGTGATGCTGTTTGATAGCATTTTA---ACAGTAGAAATTCTTTCAAA 18387

Qy       1432 A-TTGGATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              | | || ||| ||||||||||||||||       | |  | |||| || || || |||||
Db      18388 ACTAGGGTCAGTCCTCTCAAACCCTGCCACTGCCTTATTAACTAAATTGATGTAGTATTC 18447

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
                |||||| |||||| |||||||   || |||||||  |||||||||||||||| ||||||
```

```
Db      18448 CAAATCCTTTGTTGCCATTTCAGTGATGTTCACAGCATCTTCACCAGGAGTAGTTTCCAT 18507

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              ||||        |  |     |||||| ||||  |||||||| | |||||||| |||| |
Db      18508 CTCAAGAAACCACTT----TCTTTGTTCATGAATAAGAAGCAGTTCCTCATCCGTTC--G 18561

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              ||||  ||| | || ||| |||| ||||||  ||||||||     || || || |||||
Db      18562 TTTTGCCATAACATGGCAACAATTCAGTCACATCTTCAGG-----CTCCAATTCTAATTT 18616

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
              | |||||||||||||| | ||||||| | ||| |||||||   ||||||||| | ||
Db      18617 TAATTCTCTTGCTGTTTCCAACACATCTATAGTTACTTCCTCTGCTGAAGTCTTAAGCCC 18676

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              ||| |  |||| |||||| |||||||| |||||||||||| |||||||||| || ||||
Db      18677 CTCAACATCATACATGAGAGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGATAT 18736

Qy       1791 TTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
              ||||||||||  |     || |     || || ||  | |||| ||||||||| | ||
Db      18737 TTTGACCTCCTCTCTTGAATCA---CAAATATTGACATATAGAATAGTGAATCCTCTTCA 18793

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910
              |||||||||||||||||| | ||||||||||     |||||||  |||| |  ||| ||
Db      18794 GAAGGTTTTCAATTTACTTTGCCCAGATCCAT-----CAGAGGAATCACTCTGTATGGCA 18848

Qy       1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970
              | |||||||||| ||| ||||||||    |||||||  |||||||||  ||||||  |  |
Db      18849 GCTATAGCCTTATGAAATATATTTCTTAGGTAATAAGACTTGAAAGTCAAAATTAGTTTT 18908

Qy       1971 TGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||||||   |||   ||| |||||||||
Db      18909 TGATCCATGGGCTGTGGAATGGATGTTGTG 18938
```

```
RESULT 20
ABD32953
ID   ABD32953 standard; DNA; 173564 BP.
XX
AC   ABD32953;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD20-011.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
```

```
PS   claim 16; seqid 688; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 173564 BP; 48418 A; 33739 C; 37150 G; 54257 T; 0 U; 0 Other;

  Query Match             22.9%;  Score 457.4;  DB 13;  Length 173564;
  Best Local Similarity   73.9%;  Pred. No. 1.8e-68;
  Matches  749;  Conservative    0;  Mismatches  206;  Indels   58;  Gaps   11;

Qy        974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
              ||||||||  |||||  ||  ||  ||  ||    ||||  ||||||  |||  |  ||||     |  ||
Db      77710 TAGTACAGACATACTTTAAAGTTATTGCAAGTTTTGTTCCAGACCGCTACAAAGGAAGTA 77769

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
              |  ||     ||   |    |    |||                              ||||||  ||||||||||
Db      77770 ACGTGATTAAAAGAGTAGTATGAAT------------------TTTTGGTTTCCCAGTG 77810

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
              ||||||||||||  ||  |||  ||||  ||||||  |||  ||||||||  ||||||  ||||||
Db      77811 CATATAAAAGTTATGTTTACACTATACTGTA-TCTATTAAGTGTACAATAGCATTATGTT 77869

Qy       1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
              ||||||||   |  |||||||||  |||     ||   |    ||||||||||||||  ||||
Db      77870 TAAAAAAATGC--ACCTTAATTAAAAATACTTTATTGCTAAAAAAATGCTAACAGTCATC 77927

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTT---------TTCTTAT 1264
              |||  |  ||||||||||  |||  ||||||||  |||||||  |||||          ||   |  |
Db      77928 TGAACCTTCAGTGAGTGGTACTCTTTTTGTTGGTGGAGAGTCTTGCCTTGATGTTTATGT 77987

Qy       1265 TGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-- 1322
              ||  |||||  |||  ||  ||   |  ||||||||  ||  |  ||||||||||||  |||  |||||
Db      77988 TGCTGACTTATCAGGAGTTGGTTGCTGAAGGTTGGAGTGGCTGTGGCAATTTATTAAAAG 78047

Qy       1323 ---ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
                 ||||||  |||  |  ||  |  |  ||    ||||||||||||||  |||  ||||||||||  |
Db      78048 AAGACAACAATGATGTTTACCACATTGAATGACTCTTCCTTTCGTGACAGATTTCTCTGT 78107

Qy       1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-- 1437
              ||  ||||  ||||  |||||||  |||||||    ||||||||||||||||  |||||||||
Db      78108 AGCATGTGGTGCTGTTTGATAACATTTTACTAACAGTAGAACTTCTTTCAAAATTGGAGT 78167

Qy       1438 TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCC 1497
              |  |  |||  |  |  ||  ||    |||  |      |||||  |  ||||||||||  |||||
Db      78168 CAATCCTCCTCCAGCACTCCTGCTGTTTCATTGAGCAAGTTTACATAATATTCTAAATCC 78227

Qy       1498 ATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTT 1557
              |||  ||||||||||||||||  ||||||||  |||||  |||  ||||||||||||||||
Db      78228 TTTGCTGTCATTTCAACAATGTTCACAGTATCTTCCCCAAGAGTAGATTCCATCTCAAGA 78287

Qy       1558 CCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATC 1617
                 |  |    |||  ||  ||||||||||||||  ||  |||  ||  ||||||||||||  |||
Db      78288 AACCACTT----TCTCTGTTCATCCATAAGAAGCAACTCCACAGCTGTTCAAGTTTCATC 78343

Qy       1618 ATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCT 1677
              ||||||||||||  ||  ||  |||  ||||||||      ||  ||||  ||||||  ||||||
Db      78344 ATGAGATTGCAGCCATTCATTCACATCTTCAGG-----CTCCACTACTAATTCTAGTTCT 78398

Qy       1678 CTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAG 1737
              |  |||||||||  |  |||  |||||  ||   |||  |||||  ||||||||||  ||  |||||||
Db      78399 CCTGCTGTTTCCATCACTTCTGTAGTCACTTTCTCCACTGAAGTCTTGAGCCCCTCCAAG 78458

Qy       1738 TCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA-- 1795
              ||||  ||||||||||||||  ||  ||||||||||||  ||||||||||||  ||||||||
Db      78459 TCATTCATGAGGGTTGGAGTCAACTTCTTCCAAACTCCTGTTAATATTGGTATTTTGACC 78518

Qy       1796 -CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAG 1854
              |||||||||||||||||||||||||||||||||||||||||||  |||||||  |  |  |  ||||  |||
Db      78519 TCCTCCCATGAATCATGAATGTTCTTAATGGCACCTAGAATGGTAATTTTTGTCCAGAAG 78578

Qy       1855 GTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTA 1914
              |  ||||||||||||  |  ||||  ||||      ||||  |  ||  |  |  |||  ||  ||
Db      78579 GCTTTCAATTTACTTTGCCCAGGTCCA-----ACAGAAAAATCATTATGTATGATAGCTA 78633

Qy       1915 TAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
              |||||||||  ||||||||||||     |||  ||  |||||||||  ||||||||
Db      78634 TAGCCTTATTAAATGTATTTCTT---AAATGAGACTTGAAAGTCAAAATTACT 78683


RESULT 21
ACN44286/c
ID   ACN44286 standard; DNA; 105219 BP.
XX
AC   ACN44286;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG36856.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 658; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
```

```
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 105219 BP; 30555 A; 20106 C; 20937 G; 33432 T; 0 U; 189 Other;

  Query Match             22.7%;  Score 454.8;  DB 11;  Length 105219;
  Best Local Similarity   75.0%;  Pred. No. 4.9e-68;
  Matches  780;  Conservative    0;  Mismatches  187;  Indels   73;  Gaps   14;

Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
              ||||||||||||||  ||| || |||||||||| ||||  ||| || | | || ||||||
Db     104524 TACAGGCATACCTCAGAGTTATTGTGGGTTTGATTCCCGACCCCCGCGAGAAAACAAAT- 104466

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096
                              |||||| |||||||||||| | |||||| ||||||||||||
Db     104465 --------------------AATAAAATGAGTCACACAAATTTTTTGGTTTCCCAGTGCAT 104425

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156
              ||||||||| || |||||||| ||| |||||  |||  ||||||| ||  ||||||||||
Db     104424 ATAAAAGTTATGTTTATACTATACTATAGTCCATTATATGTGCAAGAGCATTATGTCTAA 104365

Qy       1157 AA---AAACACATACCTTAATTT-TAAAATGCTTTATTACTAAAAA--ATGCTAACAATC 1210
              ||   ||  | |||  |||||||  |||||| || ||| |||||||  |||||||  |||
Db     104364 AAACCAATTATATATGTTAATTTAAAAAATGGTTCATTGCTAAAAAACATGCTAATGATC 104305

Qy       1211 ATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGA 1270
              || ||||| ||||| ||||||||||||   ||||| || || || |||  |||  |
Db     104304 ATCTGAGCCTTCAGCGAGTTGTAATC---TTGCTCGTTGAGGGCCTTGCCTTGAT----- 104253

Qy       1271 CTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACA 1325
                    || |  || ||||||||  ||||||||||  ||| |||||||||      |||
Db     104252 -------GGTGGTAGTTGCTGAAGGCTAGGGTGGCTGGTGCAATTTCTTAAAATAAGACA 104200

Qy       1326 ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTG 1385
              || ||||| |||| | |||| |||||||||| |||||   | |||||||||  |||  ||
Db     104199 GCAATGAAGTTTGCCACATCAATTGACTCTTCTTTTCACAGAGGATTTCTCTGTAGCATG 104140

Qy       1386 TGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCC 1444
              |||| || |||||||| |||||  |||||||||| |||||| |||||||| |||||||
Db     104139 TGATACTCTTTGATAGCCTTTTAATCACAGTAGAATTTCTTTCAAAATTGGAGTCAATCC 104080

Qy       1445 TCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTG 1504
              |||| || |||||     ||| |  | ||||||| || | ||||||| ||||  ||||||
Db     104079 TCTCGAATCCTGCCACTGTTTTATTAACTAAGTTTATGTTATATTCTAAATCATTTGTTG 104020

Qy       1505 TCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGA 1564
              ||||||  ||| | || ||||  |||||||||||||| ||||||||||       |
Db     104019 TCATTTGGACAGTGTTTACAGCCTCTTCACCAGGAGTATATTCCATCTCAAGAAACCATT 103960

Qy       1565 TGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGAT 1624
              |    || ||| |||||||||||||  || ||||||||||       ||  |  ||||||
Db     103959 T----TCCTTGTTCATCCATAAGAAACAACTCCTCATCTG-------TTAGTTTTGAGAT 103911

Qy       1625 TGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTG 1684
              ||||||||  ||||||  ||||||||      || |||  |||||| |||||| |||||
Db     103910 TGCAGCAAGTCAGTCACATCTTCAGG-----ATTTACTACTAATTCTAGTTCTTTTGCTA 103856

Qy       1685 TTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCA 1744
              |||| |||||| |||  ||| |||| | || |||||||||||||| ||| ||||||||||
Db     103855 TTTCAACCACACCTGCAGTTACTTCATACACTGAAGTCTTGAACCCCTCAAAGTCATCCA 103796

Qy       1745 TGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT----CC 1800
              ||| ||||||||| |||||||||| | |||||||||| || ||||| | || |     ||
Db     103795 AGAGAGTTGGAATCAACTTCTTCCATACTCCTGTTAATGTTGATATTCTTACATCCTCCC 103736

Qy       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860
              |||||||||   ||||||||  | ||| || ||| |||||||||||| || ||||||||
Db     103735 CATGAATCAGAGATGTTCTTGGTAGCATCTAGAAAGGTGAATCCTTT-CAGAAGGTTTTA 103677

Qy       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920
              |||||||||  | |||| ||||       ||||||   ||| |||||| ||||| | ||||
Db     103676 AATTTACTTTGCCCAGGTCCAT-----GAGAGGAATTACTATCAATGGCAGTTGTGGCCT 103622

Qy       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980
              |||| ||||||||| || |||||| |||||| |||||| |||||||| |||||||||
Db     103621 TATGAAATGTATTTGTTAAATAATCAGGCTTAAAAGTTAAAATTACTTCTTGATCCATGG 103562

Qy       1981 TCTGCAAAATAGATGTTGTG 2000
              ||||| ||| |||||||||
Db     103561 GCTGCAGAATGGATGTTGTG 103542


RESULT 22
ABD32602
ID   ABD32602 standard; DNA; 277616 BP.
XX
AC   ABD32602;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD7-221.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 109; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
```

```
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 277616 BP; 84510 A; 49827 C; 49846 G; 93433 T; 0 U; 0 Other;

  Query Match             22.7%;  Score 454.2;  DB 13;  Length 277616;
  Best Local Similarity   73.6%;  Pred. No. 6.4e-68;
  Matches  777;  Conservative    0;  Mismatches  213;  Indels   65;  Gaps   13;

Qy        970 TAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAAT 1029
              |||| | ||||||||||||||||||||| ||  ||||| |||||| ||||||||||||
Db     173388 TAGAGAATACAGGCATACCTTGGAGATATTGCAGGTTTAGTTCCAGACCACCACAATA-- 173445

Qy       1030 ACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCC 1089
                    ||||| |||||| |||||||||| |||||||| | |||| | |||||
Db     173446 -------------AAGTGAATATCAAAATAAAGTGAATCACACAACTTTTTTTGTTTCCC 173492

Qy       1090 AGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTA 1149
              | | |||||||| ||| || ||| |||| |||||||||| | |||||||| ||||  |||
Db     173493 ACTACATATAAAGGTTATGTTTACACTATACTGTAGTCTTTAAAGTGTGCGATAGCATTA 173552

Qy       1150 TGTCTAAAAAAACACATA-----CCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTA 1204
              |||||||||||  ||||      | | | || ||||| ||||||| |||||||||  ||
Db     173553 GGTCTAAAAAAAACCATATGCATACCTTAATTAAAAATACTTTATTGCTAAAAAATTTTA 173612

Qy       1205 ACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTAT 1264
              |||||||  | |||  | |||  |||| |||||||| |||||| ||| ||| ||  || |
Db     173613 GCAATCATCTAAGCTTGCAGCAAGTTATAATCTTTATGCTGGGGGAGGGTTTTGACTCAA 173672

Qy       1265 TGATG----------ACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGT 1314
              |  ||          |||||| |||    |||      |   | ||||||||||||  |
Db     173673 TATTGATGGCTGCCCACTGATCAGGGTGATGGTTGCTGAAGGTTGGGTGGCTGTGGCCAT 173732

Qy       1315 TTCT-----TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAG 1369
              || |     ||| |||||| ||||| |||| | |||| ||||||||||||||||  ||||
Db     173733 TTTTTTAAATAAGACAACAATGAAGTTTGCCACATCAATTGACTCTTCCTTTCAGAAAAG 173792

Qy       1370 ATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTCA 1429
              ||||||||  ||  |||| |||| |||||| | ||||||  |||||||||| ||||| |
Db     173793 ATTTCTCTGCAGCATGTGGTGCTGTTTGATGGTATTTTACCCACAGTAGAACCTCTTTCA 173852

Qy       1430 AAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATAT 1488
              |||||||| | | || ||||| ||||||||     || |  | ||||||| |||||||||
Db     173853 AAATTGGAGTGAGTCTTCTCAGACCCTGCTGCTGCTTTATCAACTAAGTTTATATAATAT 173912

Qy       1489 TCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCC 1548
              ||| | ||| ||  |||||||||||| | ||| |||  |||||||||| |||||||||
Db     173913 TCTAAGTCCTTTAGTGTCATTTCAACACTGTTCCCAGCATCTTCACCAGG-GTAGATTCC 173971

Qy       1549 ATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCA 1608
              |||||          | |||       |     |||||||| || | ||||||||||
Db     173972 ATCTC-----------TAGAAATCACTTTTGCTCATAAGAAGCAACTGCTCATCTGTTCT 174020

Qy       1609 AGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAAT 1668
              | |||||| || |||||||||||||| || |||  ||||||        ||  |||| ||||
Db     174021 ACTTTTATTATAAGATTGCAGCAATTCATTCACATCTTCA-----TTCTCTACTTATAAT 174075

Qy       1669 TCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAAC 1728
              |  ||||||||||  ||| |||| || ||  ||| ||| |||||||||||||||||||
Db     174076 TGTTGTTCTCTTGCTACTTCCACCATATTTGCAGTTACTTACTCCATTGAAGTCTTGAAA 174135

Qy       1729 CTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTAT 1788
              | ||| ||||||||  |||| ||| ||||  |||||||||||| |||||||||||| ||
Db     174136 CCCTCAAAGTCATCTGTGAGAGTTAGAATTAACTTCTTCCAAACTCCTGTTAATATTGAT 174195

Qy       1789 ATTTTGACC---TCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCT 1845
              ||||||||   ||  ||||||| ||||| ||||| ||| || || |||||    || |||
Db     174196 GTTTTGACCTCTTCTAATGAATCCTGAATTTTCTTCATGACATCTAGAATG--AAACCCT 174253

Qy       1846 TTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAA 1905
              || || ||||||| ||||||  || | ||||||||||      ||| ||  |||| |  |
Db     174254 TTTCAGAAGGTTTCCAATTTGTTTTG-CCAGATCCAT-----AAGAAGAATCACTATTTA 174307

Qy       1906 TGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTA 1965
              || ||| |||| | |||   ||||| |||||| ||||||||| ||||||| || |||  |
Db     174308 TGACAGCTATA-CATTACACAATGTGTTTCTTAAATAATAAGACTTGAAAATTAAAAGCA 174366

Qy       1966 CTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              ||| |||||||||   ||||| ||| | | |||||
Db     174367 CTCTTTGATCCATGGGCTGCAGAATGGGTATTGTG 174401


RESULT 23
AAK90524
ID   AAK90524 standard; DNA; 32190 BP.
XX
AC   AAK90524;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 4100.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001324.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
```

PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
PS Disclosure; SEQ ID NO 4100; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 32190 BP; 10714 A; 6313 C; 5965 G; 9198 T; 0 U; 0 Other;

Query Match 22.7%; Score 454; DB 4; Length 32190;
Best Local Similarity 74.1%; Pred. No. 6.5e-68;
Matches 781; Conservative 0; Mismatches 215; Indels 58; Gaps 14;

```
Qy       970 TAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAAT 1029
             ||   | |||||| ||||||| |||||| ||  | |||||||||| ||||   ||||||
Db     29296 TATTAATTACAGGTATACCTTAGAGATATTGCAGTTTTGGTTCCAGACCATTGCAATAAA 29355

Qy      1030 ACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCC 1089
               |||               |||||| ||||||| ||| ||||||  ||||||  |||||
Db     29356 GCAA--------------ATATCAAAATAAAGACAGTTACACAAACCTTTTGT-TTCCC 29399

Qy      1090 AGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTA 1149
             ||||||||||||  |    ||||| || | |||| |||  |||||||| ||||   |||
```

```
Db      29400 AGTGCATATAAAACAT----TTATATTATATTGTAATCTAGTAAGTGTGTAATACCATTA 29455

Qy       1150 TGTCT-AAAAAAACACATACCTTAATT-TTAAAATGCTTTATTACTAAAAAATGCTAACA 1207

Db      29456 TGGCTGAAAAATATACATATCTCAATTCAAAAAATATTTTAATGCTAAAAAATGCTAATG 29515

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267

Db      29516 TTCATCTGAGCCTTCTGCAAGATGTGGTCTTTTTGCTGGTGGAGGGGCTAGCCTTGATAT 29575

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316

Db      29576 TGATGGCTGCTGATTGATCTGGGTAGTGGTTGCTCAAGGTTGGGGTGGCTGTGGCAACTT 29635

Qy       1317 CTTAAAACA----ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATT 1372

Db      29636 CTTAAAATATGACAATGAAGAGTTTGCTGCAACAATTGACTTTTCTTTTCACAAAAGATT 29695

Qy       1373 TCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAA 1432

Db      29696 TATCTGTAATGTGTGATGCTGTTTGATAGCATTTTACCCACAGTAGAACTTCTTTCAAAA 29755

Qy       1433 TTGGATC-AATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCT 1491

Db      29756 CTGGAGCTAATCTTCTCAAAACCTGCCACTGCTTTATCAACTAAATTTATGTAATCTTCT 29815

Qy       1492 GAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATC 1551

Db      29816 GAATTCTTTGTTGTCATTTCAACAGTGTTCACAGCATCTTCACCAAGAGTAGATTCCATC 29875

Qy       1552 TCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGT 1611

Db      29876 TCAAGAAACCACTT----TGGTTGCTTGTCCAGAAGAAGCAACTCCTCATTCATCCAAGT 29931

Qy       1612 TTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCC 1671

Db      29932 TTTATCATGAGATTGCAGCAATTCAGTCACATCTTCAG-----AATGCCCTTGTAATTCT 29986

Qy       1672 AGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTC 1731

Db      29987 AATTCTCTTGCTATTTCTACCACATGTGCAGTGACTCCCTCCACTGAAGTATTGAACCCC 30046

Qy       1732 TCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATT 1791

Db      30047 TCAAACTCATTCATGAGAATTAGAATGAACTTCTTCCCAACTCCTGTTAATGTTAATATT 30106

Qy       1792 TTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTC 1848

Db      30107 CTGACCTCCTCCCGTGAATCACAAATGTTCTTAATGGCATCTAGAATGGTGAATCCTTTC 30166

Qy       1849 CAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGC 1908

Db      30167 CAGATGGTTTTAAATTTACTTTGCCCAGCTCCAT-----CACAGGACTCACTATCTATGT 30221

Qy       1909 CAGTTATAGCCTTATGGAATGT--ATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTAC 1966

Db      30222 CAGCAATAGCCTTATGAAATGTAAATTTCTTAAGTAATAGGCCTTAAAAGTCGAAATTAC 30281

Qy       1967 TCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000

Db      30282 TCCCTGACTCA-TGGCTGTAGAATGGATGTTGTG 30314


RESULT 24
ADC85392
ID   ADC85392 standard; DNA; 96587 BP.
XX
AC   ADC85392;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human Jak2 genomic sequence.
XX
KW   Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW   secreted; transmembrane; intracellular; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003045230-A2.
XX
PD   05-JUN-2003.
XX
PF   02-DEC-2002; 2002WO-US038582.
XX
PR   30-NOV-2001; 2001US-00997722.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-513603/48.
XX
PT   New recombinant nucleic acid comprising a nucleotide sequence of any of
PT   the carcinoma-associated (CA) genes, useful for screening for drug
PT   candidates for diagnosing or treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 178; 983pp; English.
XX
CC   The invention relates to a recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the fully defined carcinoma-
CC   associated (CA) genes from the 50 tables given in the specification. The
CC   CA proteins are secreted, transmembrane or intracellular proteins. The
CC   recombinant nucleic acids are useful for screening for drug candidates
CC   for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC   ADC85514 represent CA genes of the invention.
XX
SQ   Sequence 96587 BP; 28110 A; 17495 C; 18527 G; 32435 T; 0 U; 20 Other;

  Query Match           22.7%;  Score 453;  DB 10;  Length 96587;
  Best Local Similarity  73.1%;  Pred. No. 9.9e-68;
  Matches  764;  Conservative    0;  Mismatches  225;  Indels   56;  Gaps   12;

Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032

Db      51948 AAAATACTGACACACCTCGGAGATATTGTGGGTTTGGTTCCAGACTATCGCAGTGAAGCA 52007

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092

Db      52008 AATATTG--------------CAATAAAGCAAATCACATGAAGTTTTAGGTTTCCCAGT 52052

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152

Db      52053 GCATATAAAAGTTATGTTTAAATTATTCTGTAGTCTGTTAAGTATGCAATAATATTATGT 52112

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212

Db      52113 CCAAAAAAGTACATACCTTAATTTAAAAATACTTTATTGTCAAAAAGCACTCATAATCAT 52172

Qy       1213 TTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTG------ 1266

Db      52173 CTGAGCCTTCAGTGAGTCATAATCTTTTTGTTAGTGGAGAGTCTTGCCTTGATGTTGATG 52232

Qy       1267 ----ATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA 1322

Db      52233 GCTACTGACTGACCAGGGTGGTGGCTGCTGAAGGTTGGGTGGCTATGGCAATATCTTAAA 52292

Qy       1323 ACA--ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380

Db      52293 ATAAGACAATGAAGGTTGCTGCATCAGTGGACTCTTCCTTTCGTGAAAGATGTCTCTGTA 52352

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA--T 1438

Db      52353 GTATGCAATGCTATTTGATAGCATTTTACCCGCAGTAGAACTTCTTTGAAAATAGGAAGT 52412

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
```

```
Db      52413 TGATTCACTCAAACCCTGCCACTGCTT---TGATTAAGTTGATGTAGTATTATAAATCCT 52469

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558

Db      52470 TTGTTGTCATTTCAACAATATTCACAGCATCTTCACCAGGAATAGATTCCATCTCAAGAA 52529

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618

Db      52530 ACCACTT----TCTTTGCTCATCCATAAAAAGCAATTCCTTATATGTCCAAGTTTTATCG 52585

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678

Db      52586 TGAGATTACAGCAATTCAGTCACGTCTTCAAG-----CTCCACTTCTAATTTTAGTTCTC 52640

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738

Db      52641 TTGTTACTTCTACCACATCTGCAGTGACTTCCTTCACTGAAG-ATTAAATCCCTCAAAGA 52699

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798

Db      52700 TACCCATGAGAGCCAGAATCAGCTTCTTCCAAACTCCTGTTAATGTTGACATTTTGACCT 52759

Qy       1799 CC---CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855

Db      52760 CCTCTCATGAACTACAAATGTTTTTCATGACGTCTAGAATGGTGAACTCTTTCCATAAGT 52819

Qy       1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915

Db      52820 TTTTTTATTTACTTTGCTTAGATCTGT-----GAGAGGAAACACT----GTGACAGCCAT 52870

Qy       1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975

Db      52871 AACCTTACTAAATGTGTTTCCCAAATAATAAGATTTGAAAGT--CAAATTGCTCCTTATC 52928

Qy       1976 CATTTTCTGCAAAATAGATGTTGTG 2000

Db      52929 CATGGGCTGCAGAATGGATGTTGTG 52953
```

RESULT 25
ADA02912
ID ADA02912 standard; DNA; 96589 BP.
XX
AC ADA02912;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human JAK2 carcinoma associated gene, SEQ ID NO:1430.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
PS Claim 1; SEQ ID NO 1430; 245pp; English.

XX
CC The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 96589 BP; 28111 A; 17494 C; 18529 G; 32435 T; 0 U; 20 Other;

```
  Query Match             22.7%;  Score 453;  DB 9;  Length 96589;
  Best Local Similarity   73.1%;  Pred. No. 9.9e-68;
  Matches  764;  Conservative    0;  Mismatches  225;  Indels   56;  Gaps   12;

Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032

Db      51948 AAAATACTGACACACCTCGGAGATATTGTGGGTTTGGTTCCAGACTATCGCAGTGAAGCA 52007

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092

Db      52008 AATATTG----------CAATAAAGCAAATCACATGAAGTTTTAGGTTTCCCAGT 52052

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152

Db      52053 GCATATAAAAGTTATGTTTAAATTATTCTGTAGTCTGTTAAGTATGCAATAATATTATGT 52112

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212

Db      52113 CCAAAAAAGTACATACCTTAATTTAAAAATACTTTATTGTCAAAAAGCACTCATAATCAT 52172

Qy       1213 TTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTG------ 1266

Db      52173 CTGAGCCTTCAGTGAGTCATAATCTTTTTGTTAGTGGAGAGTCTTGCCTTGATGTTGATG 52232

Qy       1267 ----ATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA 1322

Db      52233 GCTACTGACTGACCAGGGTGGTGGCTGCTGAAGGTTGGGTGGCTATGGCAATATCTTAAA 52292

Qy       1323 ACA--ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380

Db      52293 ATAAGACAATGAAGGTTGCTGCATCAGTGGACTCTTCCTTTCGTGAAAGATGTCTCTGTA 52352

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA--T 1438

Db      52353 GTATGCAATGCTATTTGATAGCATTTTACCCGCAGTAGAACTTCTTTGAAAATAGGAAGT 52412

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498

Db      52413 TGATTCACTCAAACCCTGCCACTGCTT---TGATTAAGTTGATGTAGTATTATAAATCCT 52469

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558

Db      52470 TTGTTGTCATTTCAACAATATTCACAGCATCTTCACCAGGAATAGATTCCATCTCAAGAA 52529

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618

Db      52530 ACCACTT----TCTTTGCTCATCCATAAAAAGCAATTCCTTATATGTCCAAGTTTTATCG 52585
```

```
Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
             ||||||| ||||||| |||||| ||||||| |      || ||||| |||||  |||||||
Db      52586 TGAGATTACAGCAATTCAGTCACGTCTTCAAG-----CTCCACTTCTAATTTTAGTTCTC 52640

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
             ||| |  |||||||||||||  ||  |||||| || |||||  || || | ||| |||
Db      52641 TTGTTACTTCTACCACATCTGCAGTGACTTCCTTCACTGAAG-ATTAAATCCCTCAAAGA 52699

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
              | ||||||| |   |||||  |||||||||| |||||||||| || | ||||||||||
Db      52700 TACCCATGAGAGCCAGAATCAGCTTCTTCCAAACTCCTGTTAATGTTGACATTTTGACCT 52759

Qy       1799 CC---CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855
              ||   ||||||  |  |||||| || ||| |  || ||||||||||  ||||||| |||
Db      52760 CCTCTCATGAACTACAAATGTTTTTCATGACGTCTAGAATGGTGAACTCTTTCCATAAGT 52819

Qy       1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915
              ||||  |||||||| |   |||||  |      ||||||  |||     || |||  ||
Db      52820 TTTTTTATTTACTTTGCTTAGATCTGT-----GAGAGGAAACACT----GTGACAGCCAT 52870

Qy       1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
              | |||||   ||||| ||||   |||||||||  ||||||||   || |     | | |||
Db      52871 AACCTTACTAAATGTGTTTCCCAAATAATAAGATTTGAAAGT--CAAATTGCTCCTTATC 52928

Qy       1976 CATTTTCTGCAAAATAGATGTTGTG 2000
              |||   ||||| ||| |||||||||
Db      52929 CATGGGCTGCAGAATGGATGTTGTG 52953
```

RESULT 26
ADB72650
ID ADB72650 standard; DNA; 96589 BP.
XX
AC ADB72650;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human JAK2 gene.
XX
KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS Homo sapiens.
XX
PN WO2003008583-A2.
XX
PD 30-JAN-2003.
XX
PF 26-DEC-2001; 2001WO-US051291.
XX
PR 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2003-239337/23.
XX
PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS Claim 1; SEQ ID NO 478; 2304pp; English.
XX
CC The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human gene of the invention.
XX
SQ Sequence 96589 BP; 28111 A; 17494 C; 18529 G; 32435 T; 0 U; 20 Other;

```
  Query Match             22.7%;  Score 453;  DB 10;  Length 96589;
  Best Local Similarity   73.1%;  Pred. No. 9.9e-68;
  Matches  764;  Conservative    0;  Mismatches  225;  Indels   56;  Gaps   12;

Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032
              | | ||| | || |||| |||||||| ||||||||||||||||| || | | || | |  ||
Db      51948 AAAATACTGACACACCTCGGAGATATTGTGGGTTTGGTTCCAGACTATCGCAGTGAAGCA 52007

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092
              |||||                    |||||||  | |||||  |   ||| || ||||||||
Db      52008 AATATTG--------------CAATAAAGCAAATCACATGAAGTTTTAGGTTTCCCAGT 52052

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152
              |||||||||||||| || ||| | ||  ||||||||||||||||| |||||||| | ||||||
Db      52053 GCATATAAAAGTTATGTTTAAATTATTCTGTAGTCTGTTAAGTATGCAATAATATTATGT 52112

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212
              | |||||| |||||||||||||| ||||| |||||||    |||||    || | ||||||
Db      52113 CCAAAAAAGTACATACCTTAATTTAAAAATACTTTATTGTCAAAAAGCACTCATAATCAT 52172

Qy       1213 TTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTG------ 1266
              |||||  |||||||||  |||||||||| | |||||  |||||  |||  ||
Db      52173 CTGAGCCTTCAGTGAGTCATAATCTTTTTGTTAGTGGAGAGTCTTGCCTTGATGTTGATG 52232

Qy       1267 ----ATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA 1322
                  ||||||| | |||     ||       |    | ||||||||  |||||  | |||||||
Db      52233 GCTACTGACTGACCAGGGTGGTGGCTGCTGAAGGTTGGGTGGCTATGGCAATATCTTAAA 52292

Qy       1323 ACA--ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
              | |  ||| ||||| ||||   |||||| |||||||||||||| ||||||| |||||  ||
Db      52293 ATAAGACAATGAAGGTTGCTGCATCAGTGGACTCTTCCTTTCGTGAAAGATGTCTCTGTA 52352

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA--T 1438
              || ||  ||||| |||||||||||||||  | ||||||||||||||||||||| |||  |
Db      52353 GTATGCAATGCTATTTGATAGCATTTTACCCGCAGTAGAACTTCTTTGAAAATAGGAAGT 52412

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
                || | ||||||||||||      ||       ||||||  || || |||| | |||||
Db      52413 TGATTCACTCAAACCCTGCCACTGCTT---TGATTAAGTTGATGTAGTATTATAAATCCT 52469

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
              |||||||||||||||||| |||||||  |||||||||||| ||||||||||||||
Db      52470 TTGTTGTCATTTCAACAATATTCACAGCATCTTCACCAGGAATAGATTCCATCTCAAGAA 52529

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
                | |    |||||||||||||||| ||| ||||||| || ||| ||||||||||
Db      52530 ACCACTT----TCTTTGCTCATCCATAAAAAGCAATTCCTTATATGTCCAAGTTTTATCG 52585

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
              ||||||| ||||||| |||||| ||||||| |      || ||||| |||||  |||||||
Db      52586 TGAGATTACAGCAATTCAGTCACGTCTTCAAG-----CTCCACTTCTAATTTTAGTTCTC 52640

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
              ||| |  ||||||||||||||  ||  |||||| || |||||  || || | ||| |||
Db      52641 TTGTTACTTCTACCACATCTGCAGTGACTTCCTTCACTGAAG-ATTAAATCCCTCAAAGA 52699

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
              | |||||||  |   |||||  |||||||||| |||||||||| || | ||||||||||
Db      52700 TACCCATGAGAGCCAGAATCAGCTTCTTCCAAACTCCTGTTAATGTTGACATTTTGACCT 52759

Qy       1799 CC---CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855
              ||   ||||||  |  |||||| || ||| |  || ||||||||||  ||||||| |||
Db      52760 CCTCTCATGAACTACAAATGTTTTTCATGACGTCTAGAATGGTGAACTCTTTCCATAAGT 52819

Qy       1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915
              ||||  |||||||| |   |||||  |      ||||||  ||||     || |||  ||
Db      52820 TTTTTTATTTACTTTGCTTAGATCTGT-----GAGAGGAAACACT----GTGACAGCCAT 52870
```

```
Qy       1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
Db      52871 AACCTTACTAAATGTGTTTCCCAAATAATAAGATTTGAAAGT--CAAATTGCTCCTTATC 52928

Qy       1976 CATTTTCTGCAAAATAGATGTTGTG 2000
Db      52929 CATGGGCTGCAGAATGGATGTTGTG 52953
```

RESULT 27
ADM74507
ID ADM74507 standard; DNA; 96589 BP.
XX
AC ADM74507;
XX
DT 01-JUL-2004 (first entry)
XX
DE Human carcinoma associated (CA) nucleic acid #88.
XX
KW Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW cytostatic.
XX
OS Homo sapiens.
XX
PN US2004072154-A1.
XX
PD 15-APR-2004.
XX
PF 30-NOV-2001; 2001US-00997722.
XX
PR 22-DEC-2000; 2000US-00747377.
PR 02-MAR-2001; 2001US-00798586.
XX
PA (MORR/) MORRIS D W.
PA (ENGE/) ENGELHARD E K.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2004-328562/30.
XX
PT New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
PT lymphoma.
XX
PS Claim 1; SEQ ID NO 178; 29pp; English.
XX
CC The invention relates to new recombinant nucleic acids. The invention
CC also relates to a host cell comprising a recombinant nucleic acid or
CC expression vector, an expression vector comprising a recombinant nucleic
CC acid, a recombinant protein, a method of screening for drug candidates, a
CC method of screening for a bioactive agent capable of binding to a
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC method of screening for a bioactive agent capable of modulating the
CC activity of a CAP, a method of evaluating the effect of a candidate
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC the activity of a CAP, a method of treating carcinomas, a method of
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC propensity to carcinoma. A method of evaluating the effect of a candidate
CC carcinoma drug comprises administering the drug to a patient, removing a
CC cell sample from the patient and determining alterations in the
CC expression or activation of a gene comprising the nucleotide sequence. A
CC method of diagnosing carcinoma comprises determining the expression of
CC one or more genes comprising the nucleic acid sequence in a first tissue
CC type of a first individual and comparing the expression of the gene from
CC a second normal tissue type from the first individual or a second
CC unaffected individual, where a difference in the expression indicates
CC that the first individual has carcinoma. A method of inhibiting the
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC carcinomas comprises administering to a patient an inhibitor of CAP.
CC Neutralising the effect of a CAP comprises contacting an agent specific
CC for the CAP. The polypeptide specifically binds to the protein encoded by
CC the nucleic acid. It comprises an antibody that specifically binds to the
CC protein encoded by the nucleic acid. The nucleic acids are useful for
CC preparing a composition for diagnosing or treating carcinoma e.g.,
CC leukaemia or lymphoma. This sequence represents a human carcinoma
CC associated (CA) nucleic acid of the invention. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 96589 BP; 28111 A; 17494 C; 18529 G; 32435 T; 0 U; 20 Other;

```
  Query Match             22.7%;  Score 453;  DB 12;  Length 96589;
  Best Local Similarity   73.1%;  Pred. No. 9.9e-68;
  Matches  764;  Conservative    0;  Mismatches  225;  Indels   56;  Gaps   12;

Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032
Db      51948 AAAATACTGACACACCTCGGAGATATTGTGGGTTTGGTTCCAGACTATCGCAGTGAAGCA 52007

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092
Db      52008 AATATTG--------------CAATAAAGCAAATCACATGAAGTTTTAGGTTTCCCAGT 52052

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152
Db      52053 GCATATAAAAGTTATGTTTAAATTATTCTGTAGTCTGTTAAGTATGCAATAATATTATGT 52112

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212
Db      52113 CCAAAAAAGTACATACCTTAATTTAAAAATACTTTATTGTCAAAAAGCACTCATAATCAT 52172

Qy       1213 TTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTG------ 1266
Db      52173 CTGAGCCTTCAGTGAGTCATAATCTTTTTGTTAGTGGAGAGTCTTGCCTTGATGTTGATG 52232

Qy       1267 ----ATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA 1322
Db      52233 GCTACTGACTGACCAGGGTGGTGGCTGCTGAAGGTTGGGTGGCTATGGCAATATCTTAAA 52292

Qy       1323 ACA--ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
Db      52293 ATAAGACAATGAAGGTTGCTGCATCAGTGGACTCTTCCTTTCGTGAAAGATGTCTCTGTA 52352

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA--T 1438
Db      52353 GTATGCAATGCTATTTGATAGCATTTTACCCGCAGTAGAACTTCTTTGAAAATAGGAAGT 52412

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
Db      52413 TGATTCACTCAAACCCTGCCACTGCTT---TGATTAAGTTGATGTAGTATTATAAATCCT 52469

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
Db      52470 TTGTTGTCATTTCAACAATATTCACAGCATCTTCACCAGGAATAGATTCCATCTCAAGAA 52529

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
Db      52530 ACCACTT----TCTTTGCTCATCCATAAAAAGCAATTCCTTATATGTCCAAGTTTTATCG 52585

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
Db      52586 TGAGATTACAGCAATTCAGTCACGTCTTCAAG-----CTCCACTTCTAATTTTAGTTCTC 52640

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
Db      52641 TTGTTACTTCTACCACATCTGCAGTGACTTCCTTCACTGAAG-ATTAAATCCCTCAAAGA 52699

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
Db      52700 TACCCATGAGAGCCAGAATCAGCTTCTTCCAAACTCCTGTTAATGTTGACATTTTGACCT 52759

Qy       1799 CC---CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855
Db      52760 CCTCTCATGAACTACAAATGTTTTTCATGACGTCTAGAATGGTGAACTCTTTCCATAAGT 52819
```

```
Qy       1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915
             |||| |||||||| |   ||||| |       ||||||  ||||     || |||  ||
Db      52820 TTTTTTATTTACTTTGCTTAGATCTGT-----GAGAGGAAACACT----GTGACAGCCAT 52870

Qy       1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
             | |||||   ||||| ||||   |||||||||  ||||||||   || |    | | |||
Db      52871 AACCTTACTAAATGTGTTTCCCAAATAATAAGATTTGAAAGT--CAAATTGCTCCTTATC 52928

Qy       1976 CATTTTCTGCAAAATAGATGTTGTG 2000
              |||   ||||| ||| |||||||||
Db      52929 CATGGGCTGCAGAATGGATGTTGTG 52953
```

RESULT 28
ABL64107/c
ID ABL64107 standard; DNA; 14117 BP.
XX
AC ABL64107;
XX
DT 15-MAY-2002 (first entry)
XX
DE Breast cancer related gene sequence SEQ ID NO:2444.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
PS Claim 1; SEQ ID NO 2444; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 14117 BP; 3568 A; 3182 C; 3484 G; 3883 T; 0 U; 0 Other;

```
  Query Match             22.6%;  Score 451.6;  DB 6;  Length 14117;
  Best Local Similarity   73.0%;  Pred. No. 1.6e-67;
  Matches  766;  Conservative    0;  Mismatches  224;  Indels   60;  Gaps   12;

Qy        972 GATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATAC 1031
              ||   |||||||||||||  |||||| ||   |||  |||| | ||||||||||||||  |
Db       1108 GAAGATACAGGCATACCTCAGAGATAGTGCATGTTCAGTTCTAGACCACCACAATAAAGC 1049

Qy       1032 AAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAG 1091
              |||| |              || ||||||||||||||||||   |||||| ||  | |
Db       1048 AAATGTA--------------ACCATAAAGTGAGTCACACAAACTTTTTGGTTTTCTG 1004

Qy       1092 TGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATG 1151
              || || ||||  || |  ||| |||| ||| |||||| |||||||| | ||||  |||||
Db       1003 TGTATGTAAACATTATATTTACACTATACTATAGTCTATTAAGTGTACCATAGCATTATG 944

Qy       1152 TCTAAAAAA----ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACA 1207
                |  |||||      | ||| ||||||||||||||    ||  |    ||||||||||
Db        943 GCCTAAAAACAATGTATATAACTTAATTTTAAAAATGCTTTCTTGCTAAAAATGCTAAGG 884

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              ||||| ||||| |||||  |||| ||||||||||||||||||||||||  || | ||
Db        883 ATCATCTGAGCCTTCAGCAAGTTATAATCTTTTTGCTGGTGGAAGGTCTTGCCTCAATGT 824

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              |||  | |           ||| |     ||||| |  |||| ||||||||||| |||
Db        823 TGATGGCTGAGGACTGATCAGGGTGATGATTGCTGCAAGTTAGTGTGGCTGTGGCAATTT 764

Qy       1317 CTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              || |||     ||| |||||||| ||||  ||||||||||||||||||||||||||||
Db        763 CTGAAAATGAGACATCAGTGAAGTTTGCCGCATCAGTTGACTCTTCCTTTCATGAAAGAT 704

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
```

```
Db        703 TTCCCTGTAGCATGTGATGCTGTTTGATAGCATTTTA---ACAGTAGAAATTCTTTCAAA 647

Qy       1432 A-TTGGATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490

Db        646 ACTAGGGTCAGTCCTCTCAAACCCTGCCACTGCCTTATTAACTAAATTGATGTAGTATTC 587

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550

Db        586 CAAATCCTTTGTTGCCATTTCAGTGATGTTCACAGCATCTTCACCAGGAGTAGTTTCCAT 527

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610

Db        526 CTCAAGAAACCACTT----TCTTTGTTCATGAATAAGAAGCAGTTCCTCATCCGTTC--G 473

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670

Db        472 TTTTGCCATAACATGGCAACAATTCAGTCACATCTTCAGGC------TCCATTCTAATTT 419

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730

Db        418 TAATTCTCTTGCTGTTTCCAACACATCTATAGTTACTTCCTCTGCTGAAGTCTTAAGCCC 359

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790

Db        358 CTCAACATCATACATGAGAGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGATAT 299

Qy       1791 TTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850

Db        298 TTTGACCTCCTCTCTTGAATCA---CAAATATTGACATATAGAATAGTGAATCCTCTTCA 242

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910

Db        241 GAAGGTTTTCAATTTACTTTGCCCAGATCCAT-----CAGAGGAATCA-TCTGTATGGCA 188

Qy       1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970

Db        187 GCTATAGCCTTATGAAATATATTTCTTAGGTAATAAGACTTGAAAGTCAAAATTAGTTTT 128

Qy       1971 TGATCCATTTTCTGCAAAATAGATGTTGTG 2000

Db        127 TGATCCATGGGCTGTGGAATGGATGTTGTG 98
```

RESULT 29
AAL05284
ID AAL05284 standard; DNA; 4503 BP.
XX
AC AAL05284;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 7972.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.

```
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is
PT   used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 7972; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention
XX
```

```
SQ   Sequence 4503 BP; 1360 A; 837 C; 792 G; 1514 T; 0 U; 0 Other;

  Query Match             22.4%;  Score 447.8;  DB 4;  Length 4503;
  Best Local Similarity   71.3%;  Pred. No. 6.8e-67;
  Matches  756;  Conservative    0;  Mismatches  242;  Indels   63;  Gaps   10;

Qy        970 TAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAAT 1029
              ||   | ||||||||||| || |||||  ||    |||||| | || |||||  ||| ||
Db         99 TATTGATTACAGGCATACTTTAGAGATGTTGCAAATTTGGATTCAGACCACAGCAAAAAA 158

Qy       1030 ACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCC 1089
              ||||||                ||||||  |  |||||||||| | |||||| |||||
Db        159 GCAAATACTG--------------CAATAACGCAAGTCACACAAATGTTTTGGTTTCCC 203

Qy       1090 AGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTA 1149
              |  ||||||  |||||  || |||  ||| ||||||||||  |||||||| || ||  |||
Db        204 ACGGCATATGGAAGTTATGTTTACGCTATACTGTAGTCTAGTAAGTGTGAAAAAGCATTA 263

Qy       1150 TGTC-----------TAAAAAAACACATACCTTAATTT-TAAAATGCTTTATTACTAAAA 1197
              | ||           |  |||   ||||||||||||||   ||||  ||||||  | ||||||
Db        264 TTTCTTTTTTTTTTTTTTAAAGGGACATACCTTAATTTAGAAAACGCTTTACTGCTAAAA 323

Qy       1198 AATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTT 1257
              ||||  |||||||||| ||  | |||||   || ||||||||||||  |||||  ||||||
Db        324 AATGTTAACAATCATCTGCACCTTCAGCAGGTCGTAATCTTTTTGCCAGTGGAGGGTCTT 383

Qy       1258 TTCTTA----------TTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTG 1307
                || |           ||  ||||||||  | | |      | ||||||  || ||  ||  ||
Db        384 GCCTCAATGTTGATGGCTGCTGACTGATAGCGTGGTGATTACTGAAGGTTGGGATGCATG 443

Qy       1308 TGGCAGTTTCTTAAAACA-----ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTC 1362
              |||||  ||||||||||  |     |||  ||||  ||||  | |||||||| ||||||||||
Db        444 TGGCAATTTCTTAAAATAAGAAGACAATGAAATTTGCCACATCAGTTGATTCTTCCTTTC 503

Qy       1363 ATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACT 1422
              | |||||  ||||  ||  |||  |||||||||  |||  ||||||||||  ||||||||||
Db        504 ACGAAAGTTTTCCCTGTAGCATGTGATGCTGTTTCATAGCATTTTACCCACAGTAGAACC 563

Qy       1423 TCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAAT 1481
              ||  ||  |||||||||  |||  |  ||||||   ||  ||         ||     | |||||||  |
Db        564 TCCTTCAAAATTGGAGTCAGTTCTCTCATTCCTTGTCACAGCTTTGTCAGCTAAGTTTTT 623

Qy       1482 ATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGT 1541
              |  ||||  ||   ||||  |   |||   ||||||||  || ||||  |    ||||||  || ||  |
Db        624 GTCATATACTATATCCTTCATTGATATTTCAACTATGTTCATAACATCTTCATCATGAAT 683

Qy       1542 AGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCAT 1601
              ||||||||||||        |  |    ||||||||||||||||||||  || |||||  |
Db        684 AGATTCCATCTCAAGAAACCACTT----TCTTTGCTCATCCATAAGAAGTAACTCCTCGT 739

Qy       1602 CTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCAC 1661
              |  | |||  |||||||||||||||||||||||  |||  |  ||||||||      || |||
Db        740 CCATCCAATTTTTATCATGAGATTGCAGCAATTCAGGTACATCTTCAGG-----CTCCAC 794

Qy       1662 TTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGT 1721
              || ||||||  |||||||||||||||  |||||||||   ||  |||||||| | ||||
Db        795 TTCTAATTCTGGTTCTCTTGCTGTTTCCACCACATCTTCAGTCACTTCCTCCACTAAAGT 854

Qy       1722 CTTGAACCTCTCCAAGTCATCCATGAGGGT------TGGAATCGACTTCTTCCAAATTCC 1775
              |||||||  ||||  ||||||||||||||  |      | |||||  | ||  ||| |||  |||
Db        855 CTTGAACTTCTCAAAGTCATCCATGAGGATTCAAAATTGAATCCATTTTTTCTAAACTCC 914

Qy       1776 TGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAAT 1835
              |||||  |  |  | |      ||||   ||||||||||||| |||||  ||  ||  ||  ||||
Db        915 TGTTATTGGTATTTTGACTTCCTCTTGTGAATCATGAATGCTCTTACTGACATCTAGAAT 974

Qy       1836 GGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGAT 1895
              | |  |||   ||  || |  |||| |||||||  || |  ||||||||||     |||||  |
Db        975 GATAAATTTATTTTAATATGTTTACAATTTATTTTGCCCAGATCCAT-----CAGAGAAA 1029

Qy       1896 CCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAA 1955
              ||||  | |  ||| ||||||||||| ||||||||||| | |  |||||  |||||||
Db       1030 TCACTACCTAGGGCAGCTATAGCCTTATGAAATGTATTTCTTAAGTGATAAGACTTGAAA 1089
```

```
Qy       1956 GTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGT 1996
              ||||||| |||| ||||| ||    ||||| || ||||||
Db       1090 ATTGAAATGACTCTTTGATTCAGAGGCTGCAGAACAGATGT 1130


RESULT 30
ADD50650
ID   ADD50650 standard; DNA; 175590 BP.
XX
AC   ADD50650;
XX
DT   15-JAN-2004  (first entry)
XX
DE   BAC sequence #1 containing hCHT DNA.
XX
KW   Human; high-affinity choline transporter; hCHT; chromosome 2q12;
KW   cholinergic function; Parkinson's disease; Huntington's disease;
KW   Alzheimer's disease; schizophrenia; dysautonomia; myasthenia gravis;
KW   brain; cholinergic signalling; antiparkinsonian; anticonvulsant;
KW   nootropic; neuroprotective; neuroleptic; bacterial artificial chromosome;
KW   BAC; ds.
XX
OS   Homo sapiens.
XX
PN   US2003114399-A1.
XX
PD   19-JUN-2003.
XX
PF   23-JUL-2001; 2001US-00911077.
XX
PR   23-JUL-2001; 2001US-00911077.
XX
PA   (BLAK/) BLAKELY R D.
PA   (APPA/) APPARSUNDARAM S.
PA   (FERG/) FERGUSON S.
XX
PI   Blakely RD,  Apparsundaram S,  Ferguson S;
XX
DR   WPI; 2003-810914/76.
XX
PT   Novel isolated polynucleotide encoding human or mouse high affinity
PT   choline transporter polypeptide, useful in gene therapy to increase
PT   cholinergic function in a cell of a patient suffering from Alzheimer's
PT   disease.
XX
PS   Example 1; SEQ ID NO 13; 74pp; English.
XX
CC   The present invention relates to the isolation of polynucleotide
CC   sequences encoding human and mouse high-affinity choline transporter
CC   (hCHT and mCHT respectively), and the proteins they encode. The gene
CC   encoding hCHT is located on chromosome 2q12. The polynucleotide sequence
CC   encoding hCHT, is useful for expressing hCHT recombinantly. The hCHT
CC   polynucleotide sequence when delivered to a cell, increases cholinergic
CC   function in the cell that is in a patient having Parkinson's disease,
CC   Huntington's disease, Alzheimer's disease, schizophrenia, dysautonomia or
CC   myasthenia gravis. The hCHT antibody is useful for controlling
CC   transporter CHT proteins to the brain, and for treating the above
CC   mentioned diseases. The antibody is also useful for diagnosing the above
CC   mentioned disorders and to detect the influence of cholinergic
CC   signalling. The present sequence represents a bacterial artificial
CC   chromosome (BAC) sequence containing hCHT DNA. Note: The sequence data
CC   for this patent was obtained in electronic format directly from the USPTO
CC   web site at seqdata.uspto.gov.
XX
SQ   Sequence 175590 BP; 55029 A; 33099 C; 32863 G; 53896 T; 0 U; 703 Other;

  Query Match             22.3%;  Score 446.8;  DB 10;  Length 175590;
  Best Local Similarity   73.3%;  Pred. No. 1.1e-66;
  Matches  771;  Conservative    0;  Mismatches  217;  Indels   64;  Gaps   13;

Qy        970 TAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAAT 1029
              || | | || ||| ||| ||||||||||| || |||||||||||||| | |||| |||||

Db     133582 TACAAAATATAGGGATAGCTTGGAGATATTGGGGGTTTGGTTCCAGAGCACCTCAATA-- 133639

Qy       1030 ACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCC 1089
                          ||| | || ||||||||||||||||||||||  | |||||| |||||
Db     133640 -------------AAGCGAATGTCACAATAAAGTGAGTCACACACATTTTTTGGTTTCCC 133686

Qy       1090 AGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTA 1149
              |||||||| | ||||| ||| || |||| |||||||| | ||||||||| | | |  ||
Db     133687 AGTGCATACAGAAGTTGTGCCTACACTATACTGTAGTATATTAAGTGTGAAGTGGCATTC 133746

Qy       1150 TGTCT--AAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACA 1207
                || |   ||||   ||||||| || ||  ||||| ||||||| | ||||||||||||
Db     133747 AGTTTTGTAAAATTTACATACCCTAGTTAAAAAATCCTTTATTGCCAAAAAATGCTAATG 133806

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              ||||| ||||| || |||||||   ||||| ||||| |||||| |||||||  || | |||
Db     133807 ATCATCTGAGCCTTTAGTGAGTCCCAATCTCTTTGCAGGTGGAGGGTCTTACCTCAGTGT 133866

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              | | | |           ||| |   | |||| |  || ||||||||||| | |||
Db     133867 TAATGGCTGCTGACTGATCAGGGTGGTGGTTGCTACAAGTTGGGGTGGCTGTGGGAATTT 133926

Qy       1317 CT-----TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              ||      ||| |||||| ||||| || |   ||    |||||||||||||  |||| ||
Db     133927 CTGAAAATAAGACAACAATGAAGTTTACTGCATTGACTGACTCTTCCTTTCGGGAAATAT 133986

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
              || | | | |  |||||  | |||||||||||||||     |  |||||||||||| |||
Db     133987 TTTTATGTGGCATGTGATATTGTTTGATAGCATTTTA---CCGATAGAACTTCTTTAAAA 134043

Qy       1432 ATTGG-ATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              ||||| |||| || ||||||| ||| ||      || |  |  |||||| || ||||||||
Db     134044 ATTGGAATCAGTCTTCTCAAATCCTACTGCTACTTTATCAAGTAAGTTTATGTAATATTC 134103

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
              | |||   ||||||| |||||||||||| |||||||  |||||| ||||| || ||| |||
Db     134104 TTAAT-TTTTGTTGTTATTTCAACAATGTTCACAGAATCTTCAACAGGAATAAATTTCAT 134162

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              |||||       |  |      |||||||||||||||||||  |  ||||||| ||||||||
Db     134163 CTCATGAAACCACTTCA----TTTGCTCATCCATAAGAAGCATCTCCTCATTTGTTCAAG 134218

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              |||||||| |||| |||| |||| |||||| |||||||      || ||||| ||||||
Db     134219 TTTTATCACGAGACTGCAACAATTCAGTCACATCTTCAGG-----CTCCACTTCTAATTC 134273

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
              |||||||||||||||| |||| |||||        |||||   | |||||||||| |
Db     134274 TAGTTCTCTTGCTGTTTCCACCATATCTG-------CTCCTCTGCTAAAGTCTTGAACTT 134326

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              ||| |||||||||| |||||||||| | |||||||||||||||||| | || ||||
Db     134327 CTCAAAGTCATCCATAAGGGTTGGAATCAATTTCTTCCAAATTCCTGTTAGTGTTGATAT 134386

Qy       1791 TTTGAC---CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
              ||| ||   ||| ||||||||| |||| ||| |||  ||| || ||||||  ||||||||
Db     134387 TTTTACCTTCTCTCATGAATCACGAATATTCCTAACAGCATCTAGAATGGCAAATCCTTT 134446

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              ||| |||| |||||| |||||| | |||| | |||     |||||||   ||  || |||
Db     134447 CCAGAAGGCTTTCAACTTACTTTGCCCAGCTTCAT-----CAGAGGAATTACAATCTATG 134501

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
              ||    ||||| ||   |||||||||||| | |  |||  |||||||||  ||||||||
Db     134502 GCA--ACTAGCCCTACAAAATGTATTTCTTAAGTGTTAAAACTTGAAAGTCAAAATTACT 134559

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGT 1999
              | |||||||||   || || |||  |||||||
Db     134560 CATTGATCCATGGACTACAGAATGAATGTTGT 134591


RESULT 31
ADV77908
```

```
ID   ADV77908 standard; DNA; 175590 BP.
XX
AC   ADV77908;
XX
DT   24-FEB-2005  (first entry)
XX
DE   Human BAC clone RP11-114017 on chromosome 10 genomic DNA.
XX
KW   Choline transporter; neuromuscular disorder;
KW   autonomic nervous system disorder; central nervous system disorder;
KW   Parkinsons disease; Huntingtons chorea; genetic disorder;
KW   Alzheimers disease; degeneration; neurological disease; schizophrenia;
KW   psychiatric disease; myasthenia gravis; immune disorder; gene therapy;
KW   cns-gen.; nootropic; neuroprotective; antiparkinsonian; anticonvulsant;
KW   muscular-gen.; neuroleptic; DNA purification; BAC;
KW   bacterial artificial chromosome; RP11-114017; chromosome 10; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2004248838-A1.
XX
PD   09-DEC-2004.
XX
PF   01-DEC-2003; 2003US-00724806.
XX
PR   23-JUL-2001; 2001US-00911077.
XX
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Blakely RD,  Apparsundaram S,  Ferguson S;
XX
DR   WPI; 2005-020586/02.
DR   GENBANK; AC009933.
XX
PT   New human high affinity choline transporter (CHT) cDNA, useful for
PT   treating a neuromuscular, autonomic or central nervous system disorder,
PT   including Parkinson's disease, Huntington's disease, Alzheimer's,
PT   schizophrenia.
XX
PS   Example 1; SEQ ID NO 13; 73pp; English.
XX
CC   The present invention provides polynucleotides encoding novel high
CC   affinity choline transporters (CHTs), methods for their use in screening
CC   and therapy. The invention is useful for treating neuromuscular
CC   disorders, autonomic or central nervous system disorders such as
CC   Parkinsons disease, Huntingtons disease, Alzheimers disease,
CC   schizophrenia, dysautonomia and myasthenia gravis (MG). The invention is
CC   also useful in gene therapy. The present sequence is human BAC (bacterial
CC   artificial chromosome) clone RP11-114017 on chromosome 10 genomic DNA.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the US
CC   patent office at seqdata.uspto.gov/sequence.html?DocID=US20040248838.
XX
SQ   Sequence 175590 BP; 55029 A; 33099 C; 32863 G; 53896 T; 0 U; 703 Other;

  Query Match            22.3%;  Score 446.8;  DB 14;  Length 175590;
  Best Local Similarity   73.3%;  Pred. No. 1.1e-66;
  Matches  771;  Conservative    0;  Mismatches  217;  Indels   64;  Gaps   13;

Qy        970 TAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAAT 1029
              || | | || ||| ||| |||||||||| || |||||||||||||| | |||| |||||
Db     133582 TACAAAATATAGGGATAGCTTGGAGATATTGGGGGTTTGGTTCCAGAGCACCTCAATA-- 133639

Qy       1030 ACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCC 1089
                      ||| | || ||||||||||||||||||||||  | |||||| |||||
Db     133640 ------------AAGCGAATGTCACAATAAAGTGAGTCACACACATTTTTTGGTTTCCC 133686

Qy       1090 AGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTA 1149
              |||||||| | ||||| ||| || |||| |||||||| | |||||||| | | |   ||
Db     133687 AGTGCATACAGAAGTTGTGCCTACACTATACTGTAGTATATTAAGTGTGAAGTGGCATTC 133746

Qy       1150 TGTCT--AAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACA 1207
                || |   ||||   ||||||| || ||  ||||| ||||||| | ||||||||||||

Db     133747 AGTTTTGTAAAATTTACATACCCTAGTTAAAAAATCCTTTATTGCCAAAAAATGCTAATG 133806

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              ||||| ||||| || ||||||| |||||| ||||| |||||| ||||||  || | ||
Db     133807 ATCATCTGAGCCTTTAGTGAGTCCCAATCTCTTTGCAGGTGGAGGGTCTTACCTCAGTGT 133866

Qy       1268 TGACTGAT----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              | | | |            ||| |   | |||| |  || |||||||||||| | |||
Db     133867 TAATGGCTGCTGACTGATCAGGGTGGTGGTTGCTACAAGTTGGGGTGGCTGTGGGAATTT 133926

Qy       1317 CT-----TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              ||     ||| |||||| ||||| || |   ||    ||||||||||||||  |||| ||
Db     133927 CTGAAAATAAGACAACAATGAAGTTTACTGCATTGACTGACTCTTCCTTTCGGGAAATAT 133986

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
              || | | |  |||||  | ||||||||||||||    | ||||||||||| |||
Db     133987 TTTTATGTGGCATGTGATATTGTTTGATAGCATTTTA---CCGATAGAACTTCTTTAAAA 134043

Qy       1432 ATTGG-ATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              ||||| |||| || |||||||| ||| ||      || |  |  |||||| || ||||||||
Db     134044 ATTGGAATCAGTCTTCTCAAATCCTACTGCTACTTTATCAAGTAAGTTTATGTAATATTC 134103

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
              | |||   |||||| |||||||||| |||||||  |||||| ||||| || ||| |||
Db     134104 TTAAT-TTTTGTTGTTATTTCAACAATGTTCACAGAATCTTCAACAGGAATAAATTTCAT 134162

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              |||||      | |      |||||||||||||||||| |  ||||||| ||||||||
Db     134163 CTCATGAAACCACTTCA----TTTGCTCATCCATAAGAAGCATCTCCTCATTTGTTCAAG 134218

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              |||||||| |||| |||| |||| |||||  ||||||||     || ||||| ||||||
Db     134219 TTTTATCACGAGACTGCAACAATTCAGTCACATCTTCAGG-----CTCCACTTCTAATTC 134273

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
              |||||||||||||||| |||| |||||       |||||   | |||||||||| |
Db     134274 TAGTTCTCTTGCTGTTTCCACCATATCTG-------CTCCTCTGCTAAAGTCTTGAACTT 134326

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              ||| |||||||||||| |||||||||||| | ||||||||||||||||||| | || ||||
Db     134327 CTCAAAGTCATCCATAAGGGTTGGAATCAATTTCTTCCAAATTCCTGTTAGTGTTGATAT 134386

Qy       1791 TTTGAC---CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
              ||| ||   ||| ||||||||| |||| ||| |||  ||| || |||||  ||||||||
Db     134387 TTTTACCTTCTCTCATGAATCACGAATATTCCTAACAGCATCTAGAATGGCAAATCCTTT 134446

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              ||| |||| |||||| |||||| | |||| | |||     |||||||   ||  || |||
Db     134447 CCAGAAGGCTTTCAACTTACTTTGCCCAGCTTCAT-----CAGAGGAATTACAATCTATG 134501

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
                ||    ||||| ||   |||||||||||| | |  |||  |||||||||  ||||||||
Db     134502 GCA--ACTAGCCCTACAAAATGTATTTCTTAAGTGTTAAAACTTGAAAGTCAAAATTACT 134559

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGT 1999
              | |||||||||   || || |||  |||||||
Db     134560 CATTGATCCATGGACTACAGAATGAATGTTGT 134591


RESULT 32
ABQ88179
ID   ABQ88179 standard; cDNA; 160771 BP.
XX
AC   ABQ88179;
XX
DT   18-SEP-2002  (first entry)
XX
DE   Human osteoblast differentiation related cDNA SEQ ID NO 86.
XX
KW   Human; osteoblast; stem cell differentiation; bone tissue deposition;
KW   osteoporosis; osteopathic; ss.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200250301-A2.
XX
PD   27-JUN-2002.
XX
PF   18-DEC-2001; 2001WO-US048276.
XX
PR   18-DEC-2000; 2000US-0255882P.
PR   24-APR-2001; 2001US-0285691P.
XX
PA   (GENE-) GENE LOGIC INC.
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
PI   Mertz L;
XX
DR   WPI; 2002-557663/59.
XX
PT   Use of genes and their expression profiles associated with osteoblast
PT   differentiation for screening modulators bone formation, for diagnosing
PT   or treating e.g. osteoporosis, or as markers for the differentiation
PT   process.
XX
PS   Claim 1; SEQ ID NO 86; 78pp + Sequence Listing; English.
XX
CC   The invention relates to genes and their expression profiles are used
CC   for: (a) screening modulators of precursor stem cell differentiation into
CC   osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
CC   deposition of bone tissue, abnormal rate of osteoblast formation or
CC   osteoporosis; or (c) treating or monitoring treatment of the conditions
CC   cited in (b), or monitoring the progression of bone tissue deposition.
CC   Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC   osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
CC   induced abnormalities in bone formation or bone loss, conditions that
CC   involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
CC   skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
CC   or fibrous dysplasia. The present sequence is that of an osteoblast
CC   differentiation associated cDNA marker of the invention. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 160771 BP; 44776 A; 35622 C; 34804 G; 45569 T; 0 U; 0 Other;

  Query Match             22.2%;  Score 444;  DB 6;  Length 160771;
  Best Local Similarity   73.2%;  Pred. No. 3.4e-66;
  Matches  770;  Conservative    0;  Mismatches  215;  Indels   67;  Gaps   13;

Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
              ||||||||||||  |||||| ||| | ||| |||||| |||||  || |||  || |||
Db      67072 TACAGGCATACCTCAGAGATATTGTAGATTTAGTTCCAGACCACTGCAGTAAAGCACATA 67131

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096
                        | |||| ||   |||||||||  | ||||   ||||||||||||
Db      67132 G---------------CTCAATGAATCAAGTCACACATTTTTTTTAATTTCCCAGTGCAT 67176

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156
              ||||||||| |  |||  ||| ||| |||||| |||| |||| |||||  | |||||||
Db      67177 ATAAAAGTTATATTTACCCTATACTATAGTCTATTAAATGTGAAATAGCATAATGTCTAT 67236

Qy       1157 AAAAAC-----ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCA 1211
              ||||||     |||||||||| ||| |||||  |||||| |||||||||||||  ||||
Db      67237 AAAAACAATGTACATACCTTATTTTAAAAATATTTTATTGCTAAAAAATGCTAATGATCA 67296

Qy       1212 TTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGAC 1271
              | ||||| ||  |||||| |||||||||||||||||||||| ||||||  |||  || |||
Db      67297 TCTGAGCCTTTGGTGAGTAGTAATCTTTTTGCTGGTGGAGGGTCTTGCCTTGATGTTGAT 67356

Qy       1272 TGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320
                | |           |||     | |||||||| || ||||||||| |||| |||||||
Db      67357 GGCTGCTGACTTATCAGGGCAGTGGTTGCTGAAGGTTGGGGTGGCTGGGGCAATTTCTTA 67416

Qy       1321 AA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCT 1375
              ||     |||||| ||||| |  |   ||   || | |||| ||||||||||||||||||
Db      67417 AAATAAGACAACAATGAAGTTCACTGCATTGATTTAATCTTGCTTTCATGAAAGATTTCT 67476

Qy       1376 CTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTG 1435
              ||  ||  |  |||||| ||||| | ||||||| ||||||||||||||||| |||||||
Db      67477 CTGGAGCATAAGATGCTGTTTGACAACATTTTACCCACAGTAGAACTTCTTTCAAAATTG 67536

Qy       1436 GATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAAT 1495
              |   |   ||||||||||||| |||        | ||||||| || ||||| ||| |||
Db      67537 GGGTCA-GTACTCAAACCCTGCCCTG------TCAACTAAGTTTATGTAATACTCTAAAT 67589

Qy       1496 CCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCAT 1555
              || ||||||| ||| ||||| | |||| ||  | ||||| |||||||||  |   ||
Db      67590 CCTTTGTTGT-ATTCCAACAGTGTTCAGAGCATGTTCACAAGGAGTAGAAT---TCCATC 67645

Qy       1556 TTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAAT-----TCCTCATCTGTTCAAG 1610
              |   ||      || |||||||||| |||||| ||       |||||  |||||  |||
Db      67646 TCAAGGAACCACTCTCATTGCTCATCCACAAGAAGCAACTCAACTCCTCGTCTGTTAAAG 67705

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              |||  ||||||||||  |||  |||  |||||   |||||||       |||||||||||||  |
Db      67706 TTTGATCATGAGATCACAGAAATTCAGTCCCATCTTCAG-----TCTTCACTTTTAATCC 67760

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
              ||||||||||  |  |||| | ||||| ||  ||| |||||||||  | |||  ||||||||
Db      67761 TAGTTCTCTTGTTATTTCCATCACATGTGCAGTTACTTCCTCCACTAAAGACTTGAACCC 67820

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              || |||| |||||||||  || ||||| |||||||||||| |||||  |||| || ||||
Db      67821 TTCAAAGTTATCCATGAGAATTAGAATCAACTTCTTCCAAACTCCTGGTAATGTTGATAT 67880

Qy       1791 TTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
              |||||   |||||||||||||| |||||| |||||||||| || ||||||||||||||||
Db      67881 TTTGACCTCCTCCCATGAATCACGAATGTACTTAATGGCATCTAGAATGGTGAATCCTTT 67940

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              ||| ||| ||| | ||| |||| | |||||||||     |||| |||  |||| |  |||
Db      67941 CCAGAAGATTTCCCATTCACTTTGCCCAGATCCA-----CCAGGGGACTCACTATTTATG 67995

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
              ||| || ||  |||  |||||||||||| ||||||  | |||||||||  ||||||||
Db      67996 GCAGCTACAGGGTTACAAAATGTATTTCTTAAATAAT--GACTTGAAAGTCAAAATTACT 68053

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGT 1999
              || ||| ||    | ||| |||||||||| |
Db      68054 GCTCGATTCACAGGCAGCAGAATAGATGTTAT 68085


RESULT 33
AAK90584
ID   AAK90584 standard; DNA; 4320 BP.
XX
AC   AAK90584;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 4160.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001324.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
```

PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
PS Disclosure; SEQ ID NO 4160; 986pp; English.
XX

```
CC   The present invention provides the protein and coding sequences of a
CC   number of human digestive system antigens. These can be used in the
CC   diagnosis, treatment and prevention of digestive system disorders,
CC   including cancer, Meckel's diverticulum, bacterial or parasitic
CC   infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC   ulcerative colitis. The present sequence is a genomic DNA fragment
CC   encoding a digestive system antigen of the invention
XX
SQ   Sequence 4320 BP; 1282 A; 853 C; 822 G; 1363 T; 0 U; 0 Other;

  Query Match             22.2%;  Score 443.6;  DB 4;  Length 4320;
  Best Local Similarity   74.2%;  Pred. No. 3.5e-66;
  Matches  739;  Conservative    0;  Mismatches  204;  Indels   53;  Gaps   12;

Qy       1040 AAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATA 1099
Db        372 ATGAAGCGAGTATCACAATAGAACAAATCACATGAATTTTTTGGTTTCCCAGTGCATATA 431

Qy       1100 AAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAA 1159
Db        432 AAAGTTATGTTTACACTACACTGTGGTCTATTAAATGTGCAAGAACATTATGTCTGAAGA 491

Qy       1160 A---ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGA 1216
Db        492 ATATATACACACCTTAATTTTAAAATGTTGCTT---AAAAGATGCTAACCAATAGTCAGA 548

Qy       1217 GCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTG--- 1273
Db        549 GCTTTTAGTGAGTCGT-ATCTTTTTGCTGGTGGAAGGTCTTGCCTCGATGTTGATGGCTG 607

Qy       1274 --------ATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA----- 1320
Db        608 GTGACTGATTAGGGTGGTGACTGGTGAAAGTTGTGATGGCTGTGGTAATTTCTTAAAAAT 667

Qy       1321 ----------AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGA 1370
Db        668 ACAATTTCTTAAAACACAATGAACATTGCCACATTGATTGACTCTTCCTTTTATGAAAGA 727

Qy       1371 TTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAA 1430
Db        728 TTTCTCTGTAGCCTGCAATGTTATTTGATAGCATTTTGCTCAGAGTAGAACTTCTTTCAA 787

Qy       1431 AATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATT 1489
Db        788 AATGGGAGACAGTCCTCTAAAACCCTGCTGCTACTTTATCAACTAAGTTTATATAGTTTG 847

Qy       1490 CTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCA 1549
Db        848 CTAAATCCTTTGGTATCATTTCAACAATGTTCAGAGCATCTTAGCCAAGAGTAGAACCCA 907

Qy       1550 TCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAA 1609
Db        908 TCTCAAGAAACTACTT----TCTTTGCTGCTCCATA--AAGCGACTCCTCAACCATTAAA 961

Qy       1610 GTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTC-ACTTTTAAT 1668
Db        962 GTTTTATCATGAAATTGCAGCAATTCAGTCACATCTTCAGGCTCGACTTCTATTAAGAAT 1021

Qy       1669 TCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCC----TCCATTGAAGTCTT 1724
Db       1022 TCTAGTTCTCTTGCTATTTCCATTACATCTGGAGTTACTTACTTTTTCCATTTAAGTCTT 1081

Qy       1725 GAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATAT 1784
Db       1082 GAACCCCTCAAAGTCATCCATGAGGATTGGAATTAACTTCTTCCAAACTCCTGTTCATGT 1141

Qy       1785 TTATATTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCC 1844
Db       1142 TGATATTTTGATCTTCCATGAATCATGAATATTCTTAATGGCATCTAGAATGGTGAATCC 1201

Qy       1845 TTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCA 1904
Db       1202 TTT---ACAGGTTTTCAATTTACTTTGCCCAGCTTCAT-----CAGAGGAATCACTATCT 1253

Qy       1905 ATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATT 1964
Db       1254 GTGGCAGCTACAGCCTTACAAAATGTATTTCTCGATTAATAAGACTTGATAGTTGAAATA 1313

Qy       1965 ACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
Db       1314 ACTCCTTGATCCATGGGCTGCAGAACAGCTGTTGTG 1349


RESULT 34
AAK90585
ID   AAK90585 standard; DNA; 4320 BP.
XX
AC   AAK90585;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 4161.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001324.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
```

PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
PS Disclosure; SEQ ID NO 4161; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 4320 BP; 1282 A; 854 C; 822 G; 1362 T; 0 U; 0 Other;

Query Match 22.2%; Score 443.6; DB 4; Length 4320;
Best Local Similarity 74.2%; Pred. No. 3.5e-66;
Matches 739; Conservative 0; Mismatches 204; Indels 53; Gaps 12;

```
Qy       1040 AAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATA 1099
             | |||| |  |||||||||| |   | |||||  | | ||||||  ||||||||||||||| 
Db        372 ATGAAGCGAGTATCACAATAGAACAAATCACATGAATTTTTTGGTTTCCCAGTGCATATA 431

Qy       1100 AAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAA 1159
             ||||||  || ||| |||||||||| |||| |||| |||||||  |   |||||||| || | 
Db        432 AAAGTTATGTTTACACTACACTGTGGTCTATTAAATGTGCAAGAACATTATGTCTGAAGA 491

Qy       1160 A---ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGA 1216
              |   | ||| |||||||||||||||| |   |    ||| |     | ||||  |  ||  
Db        492 ATATATACACACCTTAATTTTAAAATGTTGCTT---AAAAGATGCTAACCAATAGTCAGA 548

Qy       1217 GCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTG--- 1273
             || || |||||| || ||||||||||||||||||||  ||   || |||  |
Db        549 GCTTTTAGTGAGTCGT-ATCTTTTTGCTGGTGGAAGGTCTTGCCTCGATGTTGATGGCTG 607

Qy       1274 --------ATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA----- 1320
                      | ||| |      || ||||  ||  | |||||||| | |||||||
Db        608 GTGACTGATTAGGGTGGTGACTGGTGAAAGTTGTGATGGCTGTGGTAATTTCTTAAAAAT 667

Qy       1321 ----------AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGA 1370
                        |||  ||| |||| ||||| | ||   |||||||||||||| ||||||||
Db        668 ACAATTTCTTAAAACACAATGAACATTGCCACATTGATTGACTCTTCCTTTTATGAAAGA 727
```

```
Qy       1371 TTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAA 1430
             ||||||| |||  ||  ||| | ||||||||||||||    || ||||||||||||| ||
Db        728 TTTCTCTGTAGCCTGCAATGTTATTTGATAGCATTTTGCTCAGAGTAGAACTTCTTTCAA 787

Qy       1431 AATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATT 1489
             ||| |||  || |||||| ||||||||||      || |  | ||||||| ||||| | |
Db        788 AATGGGAGACAGTCCTCTAAAACCCTGCTGCTACTTTATCAACTAAGTTTATATAGTTTG 847

Qy       1490 CTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCA 1549
             || ||||| ||| | |||||||||||||| |||| ||  ||||  ||| |||||||  |||
Db        848 CTAAATCCTTTGGTATCATTTCAACAATGTTCAGAGCATCTTAGCCAAGAGTAGAACCCA 907

Qy       1550 TCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAA 1609
             |||||        | |    |||||||| ||||||  |||  | |||||| |  || ||
Db        908 TCTCAAGAAACTACTT----TCTTTGCTGCTCCATA--AAGCGACTCCTCAACCATTAAA 961

Qy       1610 GTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTC-ACTTTTAAT 1668
             |||||||||||| |||||||||| ||||||  |||||||||    ||||| | |   |||
Db        962 GTTTTATCATGAAATTGCAGCAATTCAGTCACATCTTCAGGCTCGACTTCTATTAAGAAT 1021

Qy       1669 TCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCC----TCCATTGAAGTCTT 1724
             || |||||||||||| |||| |  |||||||  ||| ||| |     |||||| |||||||
Db       1022 TCTAGTTCTCTTGCTATTTCCATTACATCTGGAGTTACTTACTTTTTCCATTTAAGTCTT 1081

Qy       1725 GAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATAT 1784
             ||||| ||| |||||||||||||| |||||||  |||||||||||  ||||||| || |
Db       1082 GAACCCCTCAAAGTCATCCATGAGGATTGGAATTAACTTCTTCCAAACTCCTGTTCATGT 1141

Qy       1785 TTATATTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCC 1844
             | ||||||||| || |||||||||||||| |||||||||||| || ||||||||||||||
Db       1142 TGATATTTTGATCTTCCATGAATCATGAATATTCTTAATGGCATCTAGAATGGTGAATCC 1201

Qy       1845 TTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCA 1904
             |||   | |||||||||||||||| | |||| | |||     |||||||  |||| ||
Db       1202 TTT---ACAGGTTTTCAATTTACTTTGCCCAGCTTCAT-----CAGAGGAATCACTATCT 1253

Qy       1905 ATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATT 1964
              || ||| || |||||||   |||||||||||  | |||||| ||||| |||||||||
Db       1254 GTGGCAGCTACAGCCTTACAAAATGTATTTCTCGATTAATAAGACTTGATAGTTGAAATA 1313

Qy       1965 ACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
             |||||||||||||   ||||| || || |||||||
Db       1314 ACTCCTTGATCCATGGGCTGCAGAACAGCTGTTGTG 1349
```

RESULT 35
AEB68407/c
ID AEB68407 standard; DNA; 32767 BP.
XX
AC AEB68407;
XX
DT 06-OCT-2005 (first entry)
XX
DE Human glucocorticoid receptor nucleotide sequence SEQ ID NO:25.
XX
KW antisense oligonucleotide; glucocorticoid receptor; antidiabetic;
KW anorectic; diabetes; obesity; syndrome X; hyperglycemia; hyperlipidemia;
KW ds.
XX
OS Homo sapiens.
XX
PN US2005164271-A1.
XX
PD 28-JUL-2005.
XX
PF 20-JAN-2005; 2005US-00039629.
XX
PR 20-JAN-2004; 2004US-0538173P.
PR 03-MAR-2004; 2004US-0550191P.
XX
PA (BHAN/) BHANOT S.
PA (DOBI/) DOBIE K W.
PA (FREI/) FREIER S M.
PA (DEAN/) DEAN N M.
PA (BENN/) BENNETT C F.
XX
PI Bhanot S, Dobie KW, Freier SM, Dean NM, Bennett CF;
XX
DR WPI; 2005-521414/53.
XX
PT New antisense compound targeted to a nucleic acid molecule encoding
PT glucocorticoid receptor, useful for modulating the expression of
PT glucocorticoid receptor, or for treating or diagnosing, e.g. diabetes or
PT obesity.
XX
PS Example 12; SEQ ID NO 25; 129pp; English.
XX
CC The invention relates to an antisense compound 13-80 nucleobases in
CC length targeted to a nucleic acid molecule encoding glucocorticoid
CC receptor, where the compound is complementary to the nucleic acid
CC molecule encoding glucocorticoid receptor, and where the compound
CC inhibits the expression of glucocorticoid receptor mRNA. Also described:
CC (1) a method of inhibiting the expression of glucocorticoid receptor in a
CC cell or tissue; (2) a method of treating an animal having a disease or
CC condition associated with glucocorticoid receptor; (3) a method of
CC decreasing blood glucose levels in an animal; and (4) a method of
CC preventing or delaying the onset of an increase in blood lipid levels in
CC an animal. The compounds, compositions, and methods are useful for
CC modulating the expression of glucocorticoid receptor. They are also
CC useful for the diagnosis and treatment of diseases and conditions
CC associated with glucocorticoid receptor, e.g. diabetes (Type 2), obesity,
CC metabolic syndrome X, hyperglycemia, or hyperlipidemia. The present
CC sequence represents a human glucocorticoid receptor nucleotide sequence,
CC which is used in an example from the present invention.
XX
SQ Sequence 32767 BP; 9265 A; 5801 C; 6236 G; 11465 T; 0 U; 0 Other;

Query Match 22.1%; Score 442.6; DB 14; Length 32767;
Best Local Similarity 71.5%; Pred. No. 5.6e-66;
Matches 734; Conservative 0; Mismatches 229; Indels 64; Gaps 9;

```
Qy        974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
              || |||||||||||||  |||||| |   |||| ||||| | ||||| | |
Db      30604 TAATACAGGCATACCTCTGAGATATTTCAGGTTCGGTTCTAGACCACAATA--------- 30554

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
                  ||||| ||||  ||| ||||  ||||| || | | |||||| |||||| |
Db      30553 ---------AAGTGAATATTGCAACAAAGCAAGTCAGACGAATTTTTTGGTTTCCCACTA 30503

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
              |||||||||||| |  |||||||| |||||||||| |||||||||||||||  || ||||
Db      30502 CATATAAAAGTTATATTTATACTATACTGTAGTCTATTAAGTGTGCAATAGCATTGTGTC 30443

Qy       1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
              |  |||||   ||   |||  || |   |     |||| || ||||||||||||||||||
Db      30442 TTTAAAAATGTAT---TTACCTTAATCTTAAAATATTGCTTAAAAATGCTAACAATCATT 30386

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT----------TTTCTTA 1263
              ||| | |||||||||||||||||||||| || ||||| |||||          ||  |
Db      30385 TGAACTTTCAGTGAGTTGTAATCTTTTTCCTTGTGGAGGGTCTTGCTTCAGTGTTGATGC 30326

Qy       1264 TTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA 1323
              || ||||||||| | |    |||      |   | ||||||||||||| |||||||||
Db      30325 CTGCTGACTGATCAGAGAGGTGGTTGCTGAAGGTGGGGTGGCTGTGGCAATTTCTTAAA- 30267

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTG 1383
                | |||   ||||     || ||| ||| |||||| |||||||||| | |||
Db      30266 -------GTAGACAACAATGAAGTTTCACTTTTCTTTTCATAAAAGATTTCTTTGTAGCA 30214

Qy       1384 TGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG-ATCAAT 1442
              | |||||| |||||||||||||| |||||||||  ||||| |||| ||| || |||
Db      30213 CGAGATGCTGTTTGATAGCATTTTACCCACAGTAGAAACTCTTTCAAAACTGGAATGAAT 30154

Qy       1443 CCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGT 1502
              || |||||||| ||||     || || |  |||||| || ||||||||| ||||| ||||
Db      30153 CCACTCAAACCATGCTGCTGCTTTACCAATTAAGTTTATGTAATATTCTAAATCCTTTGT 30094
```

```
Qy       1503 TGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGA 1562
              ||| ||||||||||  |||||||  |||||||||||||||||||||||||||||      |
Db      30093 TGTTATTTCAACAAGGTTCACAGAATCTTCACCAGGAGTAGATTCCATCTCAAGAAAAAA 30034

Qy       1563 GATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAG 1622
                |    |  |   |  |||||||||| || ||||||||  |||| |||| |||||||
Db      30033 ATTTATTTGCTCATTTGTCCATAAGAAGCAACTCCTCATCCATTCACGTTTTCTCATGAG 29974

Qy       1623 ATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGC 1682
              |||||||||| | ||||  | ||| ||      || ||||| ||||||  ||||||||||
Db      29973 ATTGCAGCAATTCCGTCACATTTTCGGG-----CTCCACTTCTAATTCTAGTTCTCTTGC 29919

Qy       1683 TGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATC 1742
              ||||||||||||||||  ||  ||||||  | ||||||  ||||||| ||| || |||||
Db      29918 TGTTTCTACCACATCTGCAGTGACTTCCTAAACTGAAGTGTTGAACCCCTCAAATTCATC 29859

Qy       1743 CATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC-C 1801
               || |  |||||||||  |||||||  |||| |  ||  |||| || ||  |||||||||| |
Db      29858 TATAACTGTTGGAATCAACTTCTTTCAAACTGCTACTAATGTTGATCCTTTGACCTCCTC 29799

Qy       1802 ATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCA 1861
               ||         |  ||||||||  ||| || ||||||  ||||||||  || |||||||||
Db      29798 CTGCGAATCACAAATTCTTAATAGCATCTAGAATGGCCAATCCTTTTCAGAAGGTTTTTG 29739

Qy       1862 ATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTT 1921
              |||||||| | |||| ||||     ||||||  |  |||| || ||  ||| |||||||||
Db      29738 ATTTACTTTGCCCAGGTCCA-----CCAGAGCAATCACTGTCTATAGCAGCTATAGCCTT 29684

Qy       1922 ATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTT 1981
              |   |||||| |              |  |||||| ||||||| ||||| |||| |||
Db      29683 ACAAAATGTACT-------------GATTTGAAACTTGAAATCACTCCCTGATTCATGGG 29637

Qy       1982 CTGCAAA 1988
              || ||||
Db      29636 CTACAAA 29630
```

RESULT 36
ABX08336_06
Continuation (7 of 17) of ABX08336 from base 600001 (Human phosphodiesterase 4D (PDE4D)
WP Sequence split into 17 fragments LOCUS ABX08336 Accession Abx08336

WP	Fragment Name	Begin	End
WP	ABX08336_00	1	110000
WP	ABX08336_01	100001	210000
WP	ABX08336_02	200001	310000
WP	ABX08336_03	300001	410000
WP	ABX08336_04	400001	510000
WP	ABX08336_05	500001	610000
WP	ABX08336_06	600001	710000
WP	ABX08336_07	700001	810000
WP	ABX08336_08	800001	910000
WP	ABX08336_09	900001	1010000
WP	ABX08336_10	1000001	1110000
WP	ABX08336_11	1100001	1210000
WP	ABX08336_12	1200001	1310000
WP	ABX08336_13	1300001	1410000
WP	ABX08336_14	1400001	1510000
WP	ABX08336_15	1500001	1610000
WP	ABX08336_16	1600001	1691080

Query Match 22.1%; Score 441.6; DB 6; Length 110000;
Best Local Similarity 71.7%; Pred. No. 8.6e-66;
Matches 755; Conservative 0; Mismatches 232; Indels 66; Gaps 11;

```
Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032
              |||   ||| |||||||  |||||||  ||||||||  | ||||  ||||  ||||||  ||
Db      51179 ATATACCAGACATACCTCGGAGATATTGTGGGTTCAGATCCAGGCCACTGCAATAAAGCA 51238

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092
              |                 ||||||||||||||  |||||||||    |||||||||  |||||
Db      51239 A---------------ATATCACAATAAAGCAAGTCACACATACTTTTTGGCTTTCCAGT 51283

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152
              ||||||||||  || || ||| |||| | ||||||  ||| |     |   || || |
Db      51284 GCATATAAAACTTATGTTTACACTATATTGTAGTGTGTAA----TAGCATTGTCTCTAAG 51339

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACT-----AAAAAATGCTAACA 1207
                || ||   |||| |||||| |    |||| ||| ||| ||      ||||||||||||||
Db      51340 AAAACAATGTACATGCCTTAAGTAAGAAATACTTAATTGCTAAAAAAAAAAAATGCTAACA 51399

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              | |||| |||| |||||  |||   || | ||||||||||||||  |||||  || | ||
Db      51400 ACCATTAGAGCCTTCAGCAAGTCAGAAACATTTTGCTGGTGGAGAGTCTTGCCTCAATGT 51459

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              | |   |             || |   | ||||||||  || ||||||||||  || |||
Db      51460 TCATGACTGCTGACTGATCAAGGTGGTGGTTGCTGAAGATTGGGGTGGCTGTGACAATTT 51519

Qy       1317 CT-----TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              ||     ||| || |||||||| ||||| | |||  |||| || ||||||||||  ||
Db      51520 CTTAGAATAACACTACAGTGAAAATTGCCACATCCATTGATTCCTCCTTTCATGAACAAT 51579

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
              |||||| |||  ||  ||| | ||||| |||||||||  || ||||||||||||||| |||
Db      51580 TTCTCTGTAGCATGCAATGGTGTTTGACAGCATTTTACCCATAGTAGAACTTCTTTCAAA 51639

Qy       1432 ATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              |||||| |||||||||| |||| |||||     || |     |||||| ||||||||||
Db      51640 ATTGGAGTCAATCCTCTTAAACGCTGCTGCTGGTTTATCTACTAAGTTTATATAATATTT 51699

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
              | ||||| |||||| |||||||||||| |||||||  |||| |||||  ||||||||  |
Db      51700 TAAATCCTTTGTTGCCATTTCAACAATGTTCACAGCATCTTTACCAGAAGTAGATTCTGT 51759

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              ||||        |  |     |  ||  ||      |||||  || |||||||   ||||||
Db      51760 CTCAAGAAACCATTTTCTTTGCTTCATC-----CAAGAAACAACTCCTCATTCTTTCAAG 51814

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              |||||||| |||||||| |||     | |    |||||||       | || || ||||||
Db      51815 TTTTATCATAAGATTGCAGTAATTTCATTACAACTTCAGG-----TTCCATTTCTAATTC 51869

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
              ||||||||||  |||| | ||||||||  ||| ||| |||||       ||||||||
Db      51870 TAGTTCTCTTGCAATTTCCAGCACATCTGCAGTTACTTTCTCCA-------CTTGAACCC 51922

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              |||  |||||||||||||||||||||| | ||||||||||  || |||||||  || ||||
Db      51923 CTCAGAGTCATCCATGAGGGTTGGAATCAATTTCTTCCAAACTCTTGTTAATGTTGATAT 51982

Qy       1791 TTTGACCT---CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
              ||||||||   || ||| |||  ||| |||||||||||| || ||||||||||||| |
Db      51983 TTTGACCTATGCCTATGGATCACAAATATTCTTAATGGCATCTAGAATGGTGAATCCTGT 52042

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              ||| || |||||| | |||||| | || |||||||     |||||  |   | | || ||
Db      52043 CCAGAAAGTTTTCGACTTACTTTGCCCTGATCCAT-----CAGAGAAATAAGTGTCTATC 52097

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
                ||  |||||| |||| |||||||||||| |||||||||||||| |||||||||| |||
Db      52098 ATAGCGATAGCCCTATGAAATGTATTTCTTAAATAATAAGGCTTAAAAGTTGAAATGACT 52157

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||||   |||  || || |||  ||||||||
Db      52158 CCTTGGAACATGGGCTTCAGAATGAATGTTGTG 52190
```

RESULT 37
ADJ25985_06
Continuation (7 of 17) of ADJ25985 from base 600001 (Human phosphodiesterase 4D (PDE4D)
WP Sequence split into 17 fragments LOCUS ADJ25985 Accession Adj25985

WP	Fragment Name	Begin	End
WP	ADJ25985_00	1	110000
WP	ADJ25985_01	100001	210000
WP	ADJ25985_02	200001	310000

```
WP      ADJ25985_03         300001    410000
WP      ADJ25985_04         400001    510000
WP      ADJ25985_05         500001    610000
WP      ADJ25985_06         600001    710000
WP      ADJ25985_07         700001    810000
WP      ADJ25985_08         800001    910000
WP      ADJ25985_09         900001   1010000
WP      ADJ25985_10        1000001   1110000
WP      ADJ25985_11        1100001   1210000
WP      ADJ25985_12        1200001   1310000
WP      ADJ25985_13        1300001   1410000
WP      ADJ25985_14        1400001   1510000
WP      ADJ25985_15        1500001   1610000
WP      ADJ25985_16        1600001   1691139

  Query Match             22.1%;  Score 441.6;  DB 12;  Length 110000;
  Best Local Similarity   71.7%;  Pred. No. 8.6e-66;
  Matches  755;  Conservative    0;  Mismatches  232;  Indels   66;  Gaps   11;

Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032
              |||   ||| ||||||| ||||||| ||||||||  | ||||  ||||  ||||||  ||
Db      51238 ATATACCAGACATACCTCGGAGATATTGTGGGTTCAGATCCAGGCCACTGCAATAAAGCA 51297

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092
              |               ||||||||||||||  |||||||||     |||||||||  |||||
Db      51298 A---------------ATATCACAATAAAGCAAGTCACACATACTTTTTGGCTTTCCAGT 51342

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152
              ||||||||||  ||  ||  |||  ||||  | ||||||  |||  |     |   ||  | |
Db      51343 GCATATAAAACTTATGTTTACACTATATTGTAGTGTGTAA----TAGCATTGTCTCTAAG 51398

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACT-----AAAAAATGCTAACA 1207
                ||  ||    ||||  ||||||  |    ||||  |||  |||  ||      ||||||||||||||
Db      51399 AAAACAATGTACATGCCTTAAGTAAGAAATACTTAATTGCTAAAAAAAAAAAATGCTAACA 51458

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              | ||||  ||||  |||||   |||    ||  | ||||||||||||||  |||||   ||  | ||
Db      51459 ACCATTAGAGCCTTCAGCAAGTCAGAAACATTTTGCTGGTGGAGAGTCTTGCCTCAATGT 51518

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              | |   |           || |   | |||||||| || ||||||||||  || |||
Db      51519 TCATGACTGCTGACTGATCAAGGTGGTGGTTGCTGAAGATTGGGGTGGCTGTGACAATTT 51578

Qy       1317 CT-----TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              ||     ||| || |||||||| |||||  | |||  ||||  || ||||||||||||  ||
Db      51579 CTTAGAATAACACTACAGTGAAAATTGCCACATCCATTGATTCCTCCTTTCATGAACAAT 51638

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
              ||||||  |||   ||  |||  | |||||  |||||||||  ||  ||||||||||||||  |||
Db      51639 TTCTCTGTAGCATGCAATGGTGTTTGACAGCATTTTACCCATAGTAGAACTTCTTTCAAA 51698

Qy       1432 ATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              ||||||  ||||||||||  ||||  |||||      ||  |    |||||||  ||||||||||
Db      51699 ATTGGAGTCAATCCTCTTAAACGCTGCTGCTGGTTTATCTACTAAGTTTATATAATATTT 51758

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
              | |||||  ||||||  ||||||||||||  |||||||   ||||  |||||  |||||||||   |
Db      51759 TAAATCCTTTGTTGCCATTTCAACAATGTTCACAGCATCTTTACCAGAAGTAGATTCTGT 51818

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              ||||       |  |    |  ||  ||      |||||  ||  |||||||   ||||||
Db      51819 CTCAAGAAACCATTTTCTTTGCTTCATC-----CAAGAAACAACTCCTCATTCTTTCAAG 51873

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              |||||||||  |||||||||  |||     |  |   |||||||      |  ||  ||  ||||||
Db      51874 TTTTATCATAAGATTGCAGTAATTTCATTACAACTTCAGG-----TTCCATTTCTAATTC 51928

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
                |||||||||||   ||||  |  ||||||||   |||  |||  |||||       ||||||||
Db      51929 TAGTTCTCTTGCAATTTCCAGCACATCTGCAGTTACTTTCTCCA-------CTTGAACCC 51981

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              |||  |||||||||||||||||||||||  | ||||||||||  ||  |||||||  ||  ||||
Db      51982 CTCAGAGTCATCCATGAGGGTTGGAATCAATTTCTTCCAAACTCTTGTTAATGTTGATAT 52041

Qy       1791 TTTGACCT---CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
              ||||||||    ||  |||  ||||   |||  ||||||||||||  ||  ||||||||||||||  |
Db      52042 TTTGACCTATGCCTATGGATCACAAATATTCTTAATGGCATCTAGAATGGTGAATCCTGT 52101

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              |||  ||  ||||||  |  ||||||  |  ||  |||||||      |||||  |    |  |  ||  ||
Db      52102 CCAGAAAGTTTTCGACTTACTTTGCCCTGATCCAT-----CAGAGAAATAAGTGTCTATC 52156

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
                ||  ||||||  ||||  ||||||||||||  ||||||||||||  ||||||||||  |||
Db      52157 ATAGCGATAGCCCTATGAAATGTATTTCTTAAATAATAAGGCTTAAAAGTTGAAATGACT 52216

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||||   |||   ||  ||  |||  ||||||||
Db      52217 CCTTGGAACATGGGCTTCAGAATGAATGTTGTG 52249


RESULT 38
ADN97989_06
Continuation (7 of 17) of ADN97989 from base 600001 (Human phosphodiesterase 4D genomic
WP Sequence split into 17 fragments  LOCUS ADN97989 Accession Adn97989
WP      Fragment Name       Begin     End
WP      ADN97989_00             1    110000
WP      ADN97989_01        100001    210000
WP      ADN97989_02        200001    310000
WP      ADN97989_03        300001    410000
WP      ADN97989_04        400001    510000
WP      ADN97989_05        500001    610000
WP      ADN97989_06        600001    710000
WP      ADN97989_07        700001    810000
WP      ADN97989_08        800001    910000
WP      ADN97989_09        900001   1010000
WP      ADN97989_10       1000001   1110000
WP      ADN97989_11       1100001   1210000
WP      ADN97989_12       1200001   1310000
WP      ADN97989_13       1300001   1410000
WP      ADN97989_14       1400001   1510000
WP      ADN97989_15       1500001   1610000
WP      ADN97989_16       1600001   1691138

  Query Match             22.1%;  Score 441.6;  DB 12;  Length 110000;
  Best Local Similarity   71.7%;  Pred. No. 8.6e-66;
  Matches  755;  Conservative    0;  Mismatches  232;  Indels   66;  Gaps   11;

Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032
              |||   ||| ||||||| ||||||| ||||||||  | ||||  ||||  ||||||  ||
Db      51238 ATATACCAGACATACCTCGGAGATATTGTGGGTTCAGATCCAGGCCACTGCAATAAAGCA 51297

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092
              |               ||||||||||||||  |||||||||     |||||||||  |||||
Db      51298 A---------------ATATCACAATAAAGCAAGTCACACATACTTTTTGGCTTTCCAGT 51342

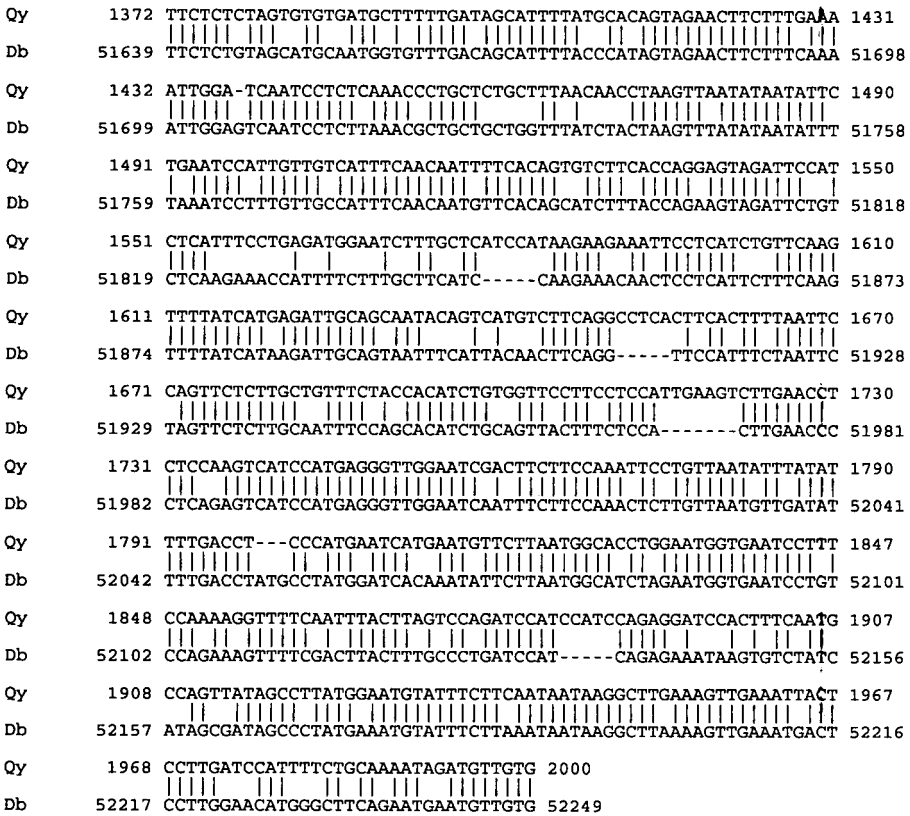
Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152
              ||||||||||  ||  ||  |||  ||||  | ||||||  |||  |     |   ||  | |
Db      51343 GCATATAAAACTTATGTTTACACTATATTGTAGTGTGTAA----TAGCATTGTCTCTAAG 51398

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACT-----AAAAAATGCTAACA 1207
                ||  ||    ||||  ||||||  |    ||||  |||  |||  ||      ||||||||||||||
Db      51399 AAAACAATGTACATGCCTTAAGTAAGAAATACTTAATTGCTAAAAAAAAAAAATGCTAACA 51458

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              | ||||  ||||  |||||   |||    ||  | ||||||||||||||  |||||   ||  | ||
Db      51459 ACCATTAGAGCCTTCAGCAAGTCAGAAACATTTTGCTGGTGGAGAGTCTTGCCTCAATGT 51518

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              | |   |           || |   | |||||||| || ||||||||||  || |||
Db      51519 TCATGACTGCTGACTGATCAAGGTGGTGGTTGCTGAAGATTGGGGTGGCTGTGACAATTT 51578

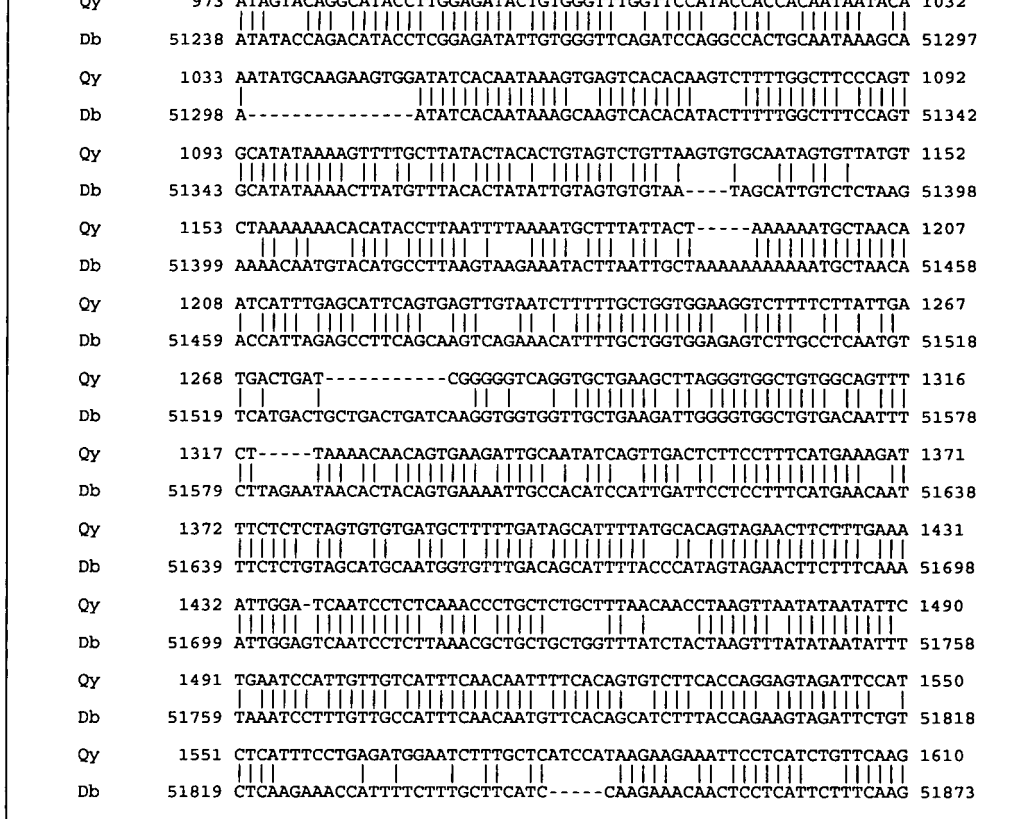
Qy       1317 CT-----TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              ||     ||| || |||||||| |||||  | |||  ||||  || ||||||||||||  ||
Db      51579 CTTAGAATAACACTACAGTGAAAATTGCCACATCCATTGATTCCTCCTTTCATGAACAAT 51638
```



RESULT 39
ADO50281_06
Continuation (7 of 17) of ADO50281 from base 600001 (Human phosphodiesterase 4D (PDE4D)
WP Sequence split into 17 fragments LOCUS ADO50281 Accession Ado50281

	Fragment Name	Begin	End
WP	ADO50281_00	1	110000
WP	ADO50281_01	100001	210000
WP	ADO50281_02	200001	310000
WP	ADO50281_03	300001	410000
WP	ADO50281_04	400001	510000
WP	ADO50281_05	500001	610000
WP	ADO50281_06	600001	710000
WP	ADO50281_07	700001	810000
WP	ADO50281_08	800001	910000
WP	ADO50281_09	900001	1010000
WP	ADO50281_10	1000001	1110000
WP	ADO50281_11	1100001	1210000
WP	ADO50281_12	1200001	1310000
WP	ADO50281_13	1300001	1410000
WP	ADO50281_14	1400001	1510000
WP	ADO50281_15	1500001	1610000
WP	ADO50281_16	1600001	1691134

Query Match 22.1%; Score 441.6; DB 12; Length 110000;
Best Local Similarity 71.7%; Pred. No. 8.6e-66;
Matches 755; Conservative 0; Mismatches 232; Indels 66; Gaps 11;



```
RESULT 40
AEB85185_06
Continuation (7 of 17) of AEB85185 from base 600001 (Human phosphodiesterase 4D gene SEQ
WP Sequence split into 17 fragments  LOCUS AEB85185 Accession Aeb85185
WP      Fragment Name       Begin     End
WP      AEB85185_00             1     110000
WP      AEB85185_01        100001     210000
WP      AEB85185_02        200001     310000
WP      AEB85185_03        300001     410000
WP      AEB85185_04        400001     510000
WP      AEB85185_05        500001     610000
WP      AEB85185_06        600001     710000
WP      AEB85185_07        700001     810000
WP      AEB85185_08        800001     910000
WP      AEB85185_09        900001    1010000
WP      AEB85185_10       1000001    1110000
WP      AEB85185_11       1100001    1210000
WP      AEB85185_12       1200001    1310000
WP      AEB85185_13       1300001    1410000
WP      AEB85185_14       1400001    1510000
WP      AEB85185_15       1500001    1610000
WP      AEB85185_16       1600001    1691140

  Query Match             22.1%;  Score 441.6;  DB 14;  Length 110000;
  Best Local Similarity   71.7%;  Pred. No. 8.6e-66;
  Matches  755;  Conservative    0;  Mismatches  232;  Indels   66;  Gaps   11;

Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032
              |||   ||| ||||||| ||||||| ||||||||  | ||||  ||||  ||||||  ||
Db      51239 ATATACCAGACATACCTCGGAGATATTGTGGGTTCAGATCCAGGCCACTGCAATAAAGCA 51298

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092
              |               ||||||||||||||  |||||||||    |||||||||| |||||
Db      51299 A---------------ATATCACAATAAAGCAAGTCACACATACTTTTTGGCTTTCCAGT 51343

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152
              |||||||||| || || ||| |||| | |||||| ||| |    |    || || |
Db      51344 GCATATAAAACTTATGTTTACACTATATTGTAGTGTGTAA----TAGCATTGTCTCTAAG 51399

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACT-----AAAAAATGCTAACA 1207
                || ||   |||| |||||| |   |||| ||| ||| ||     ||||||||||||||
Db      51400 AAAACAATGTACATGCCTTAAGTAAGAAATACTTAATTGCTAAAAAAAAAAAATGCTAACA 51459

Qy       1208 ATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGA 1267
              | |||| |||| ||||| |||   || | |||||||||||||  |||||  || | ||
Db      51460 ACCATTAGAGCCTTCAGCAAGTCAGAAACATTTTGCTGGTGGAGAGTCTTGCCTCAATGT 51519

Qy       1268 TGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT 1316
              | |    |           || |   | |||||||| || |||||||||| || |||
Db      51520 TCATGACTGCTGACTGATCAAGGTGGTGGTTGCTGAAGATTGGGGTGGCTGTGACAATTT 51579

Qy       1317 CT-----TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGAT 1371
              ||     ||| || |||||||| ||||| | |||  |||| || |||||||||||  ||
Db      51580 CTTAGAATAACACTACAGTGAAAATTGCCACATCCATTGATTCCTCCTTTCATGAACAAT 51639

Qy       1372 TTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAA 1431
              |||||| |||  ||  ||| | ||||| |||||||||  || ||||||||||||||| |||
Db      51640 TTCTCTGTAGCATGCAATGGTGTTTGACAGCATTTTACCCATAGTAGAACTTCTTTCAAA 51699

Qy       1432 ATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTC 1490
              |||||| |||||||||| |||| |||||     || |    ||||||| ||||||||||
Db      51700 ATTGGAGTCAATCCTCTTAAACGCTGCTGCTGGTTTATCTACTAAGTTTATATAATATTT 51759

Qy       1491 TGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCAT 1550
              | ||||| |||||| |||||||||||| |||||||  |||| ||||| |||||||||  |
Db      51760 TAAATCCTTTGTTGCCATTTCAACAATGTTCACAGCATCTTTACCAGAAGTAGATTCTGT 51819

Qy       1551 CTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAG 1610
              ||||       |  |    |  ||  ||       |||||  || |||||||   ||||||
Db      51820 CTCAAGAAACCATTTTCTTTGCTTCATC-----CAAGAAACAACTCCTCATTCTTTCAAG 51874

Qy       1611 TTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTC 1670
              ||||||||| ||||||||| |||    | |   |||||||      | || || ||||||
Db      51875 TTTTATCATAAGATTGCAGTAATTTCATTACAACTTCAGG-----TTCCATTTCTAATTC 51929

Qy       1671 CAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCT 1730
              ||||||||||  |||| | ||||||||  ||| ||| |||||       |||||||
Db      51930 TAGTTCTCTTGCAATTTCCAGCACATCTGCAGTTACTTTCTCCA-------CTTGAACCC 51982

Qy       1731 CTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATAT 1790
              |||  |||||||||||||||||||||| | |||||||||| || ||||||| || ||||
Db      51983 CTCAGAGTCATCCATGAGGGTTGGAATCAATTTCTTCCAAACTCTTGTTAATGTTGATAT 52042

Qy       1791 TTTGACCT---CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
              ||||||||   || ||| ||||  ||| ||||||||||||| || ||||||||||||| |
Db      52043 TTTGACCTATGCCTATGGATCACAAATATTCTTAATGGCATCTAGAATGGTGAATCCTGT 52102

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              ||| || |||||| | |||||| | || |||||||     ||||| |   | | || ||
Db      52103 CCAGAAAGTTTTCGACTTACTTTGCCCTGATCCAT-----CAGAGAAATAAGTGTCTATC 52157

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
                 || |||||| |||| |||||||||||| |||||||||||| |||||||||| |||
Db      52158 ATAGCGATAGCCCTATGAAATGTATTTCTTAAATAATAAGGCTTAAAAGTTGAAATGACT 52217

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||||   |||   || || |||  ||||||||
Db      52218 CCTTGGAACATGGGCTTCAGAATGAATGTTGTG 52250


RESULT 41
ADA02954/c
ID   ADA02954 standard; DNA; 96589 BP.
XX
AC   ADA02954;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Human NR3C1 carcinoma associated gene, SEQ ID NO:1472.
XX
KW   Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW   prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003057146-A2.
XX
PD   17-JUL-2003.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 1472; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
```

```
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed human CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 96589 BP; 28270 A; 17369 C; 18773 G; 31440 T; 0 U; 737 Other;

  Query Match             22.1%;  Score 441;  DB 9;  Length 96589;
  Best Local Similarity   71.4%;  Pred. No. 1.1e-65;
  Matches  733;  Conservative    0;  Mismatches  230;  Indels   64;  Gaps    9;

Qy        974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
              || |||||||||||||  ||||||  |   ||||  |||||  | |||||  | |
Db      58294 TAATACAGGCATACCTCTGAGATATTTCAGGTTCGGTTCTAGACCACAATA--------- 58244

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
                 |||||  ||||   |||  ||||   |||||  ||  |  |  ||||||  ||||||  |
Db      58243 ---------AAGTGAATATTGCAACAAAGCAAGTCAGACGAATTTTTTGGTTTCCCACTA 58193

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
              ||||||||||||  |  ||||||||  ||||||||||  ||||||||||||||||   ||  ||||
Db      58192 CATATAAAAGTTATATTTATACTATACTGTAGTCTATTAAGTGTGCAATAGCATTGTGTC 58133

Qy       1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
              |  |||||   ||   |||  ||  |    |     ||||  ||  ||||||||||||||||||
Db      58132 TTTAAAAATGTAT---TTACCTTAATCTTAAAATATTGCTTAAAAATGCTAACAATCATT 58076

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT----------TTTCTTA 1263
              |||  |  ||||||||||||||||||||||  ||  |||||  |||||          ||   |
Db      58075 TGAACTTTCAGTGAGTTGTAATCTTTTTCCTTGTGGAGGGTCTTGCTTCAGTGTTGATGC 58016

Qy       1264 TTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA 1323
               ||  |||||||||  |  |     |||      |    | ||||||||||||||  |||||||||
Db      58015 CTGCTGACTGATCAGAGAGGTGGTTGCTGAAGGTGGGGTGGCTGTGGCAATTTCTTAAA- 57957

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTG 1383
                   |  |||    ||||      || |||  |||  ||||||  ||||||||||  | |||
Db      57956 -------GTAGACAACAATGAAGTTTCACTTTTCTTTTCATAAAAGATTTCTTTGTAGCA 57904

Qy       1384 TGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG-ATCAAT 1442
               | ||||||  ||||||||||||||||  ||||||||||  |||||  ||||  |||  ||  |||
Db      57903 CGAGATGCTGTTTGATAGCATTTTACCCACAGTAGAAACTCTTTCAAAACTGGAATGAAT 57844

Qy       1443 CCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGT 1502
              ||  ||||||||  ||||      ||  ||  |   ||||||  ||  |||||  |||  |||||  ||||
Db      57843 CCACTCAAACCATGCTGCTGCTTTACCAATTAAGTTTATGTAATACTCTAAATCCTTTGT 57784

Qy       1503 TGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGA 1562
              |||  ||||||||||  |||||||  ||||||||||||||||||||||||||||||      |
Db      57783 TGTTATTTCAACAAGGTTCACAGAATCTTCACCAGGAGTAGATTCCATCTCAAGAAAAAA 57724

Qy       1563 GATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAG 1622
                |    |  |    |  ||||||||||  ||  ||||||||  ||||  |||||  |||||||
Db      57723 ATTTATTTGCTCATTTGTCCATAAGAAGCAACTCCTCATCCATTCACGTTTTCTCATGAG 57664

Qy       1623 ATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGC 1682
              ||||||||||  |  ||||  |  |||  ||      ||  |||||  ||||||  ||||||||||
Db      57663 ATTGCAGCAATTCCGTCACATTTTCGGG-----CTCCACTTCTAATTCTAGTTCTCTTGC 57609

Qy       1683 TGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATC 1742
              ||||||||||||||||  ||  ||||||  |  ||||||  |||||  ||  ||  |||||
Db      57608 TGTTTCTACCACATCTGCAGTGACTTCCTAAACTGAAGTGTTGAACCCCTCAAATTCATC 57549

Qy       1743 CATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC-C 1801
               ||  |  |||||||||  |||||||  ||||  |  ||  ||||  ||  ||  ||||||||||  |
Db      57548 TATAACTGTTGGAATCAACTTCTTTCAAACTGCTACTAATGTTGATCCTTTGACCTCCTC 57489

Qy       1802 ATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCA 1861
              ||         |  ||||||||  |||  ||  ||||||  ||||||||  ||  ||||||||
Db      57488 CTGCGAATCACAAATTCTTAATAGCATCTAGAATGGCCAATCCTTTTCAGAAGGTTTTTG 57429

Qy       1862 ATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTT 1921
              ||||||||  |  ||||  ||||      ||||||  |  ||||  ||  ||  |||  |||||||||
Db      57428 ATTTACTTTGCCCAGGTCCA-----CCAGAGCAATCACTGTCTATAGCAGCTATAGCCTT 57374

Qy       1922 ATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTT 1981
              |    ||||||  |              |  ||||||  |||||||  |||||  ||||  |||
Db      57373 ACAAAATGTACT-------------GATTTGAAACTTGAAATCACTCCCTGATTCATGGG 57327

Qy       1982 CTGCAAA 1988
              ||  ||||
Db      57326 CTACAAA 57320


RESULT 42
ADB72692/c
ID   ADB72692 standard; DNA; 96589 BP.
XX
AC   ADB72692;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human NR3C1 gene.
XX
KW   human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 520; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 96589 BP; 28270 A; 17369 C; 18773 G; 31440 T; 0 U; 737 Other;

  Query Match             22.1%;  Score 441;  DB 10;  Length 96589;
  Best Local Similarity   71.4%;  Pred. No. 1.1e-65;
  Matches  733;  Conservative    0;  Mismatches  230;  Indels   64;  Gaps    9;

Qy        974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
              || |||||||||||||  ||||||  |   ||||  |||||  | |||||  | |
Db      58294 TAATACAGGCATACCTCTGAGATATTTCAGGTTCGGTTCTAGACCACAATA--------- 58244
```

```
Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
                ||||| ||||  ||| ||||  ||||| || | | |||||| |||||| |
Db      58243 ---------AAGTGAATATTGCAACAAAGCAAGTCAGACGAATTTTTTGGTTTCCCACTA 58193

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
                |||||||||||| |  |||||||| |||||||||| ||||||||||||||   || ||||
Db      58192 CATATAAAAGTTATATTTATACTATACTGTAGTCTATTAAGTGTGCAATAGCATTGTGTC 58133

Qy       1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
                |  |||||   ||   |||  || |   |    |||| || ||||||||||||||||||
Db      58132 TTTAAAAATGTAT---TTACCTTAATCTTAAAATATTGCTTAAAAATGCTAACAATCATT 58076

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT----------TTTCTTA 1263
                ||| | ||||||||||||||||||||||| || ||||| |||||          ||  |
Db      58075 TGAACTTTCAGTGAGTTGTAATCTTTTTCCTTGTGGAGGGTCTTGCTTCAGTGTTGATGC 58016

Qy       1264 TTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA 1323
                || |||||||| | |    |||     |   | ||||||||||||| |||||||||
Db      58015 CTGCTGACTGATCAGAGAGGTGGTTGCTGAAGGTGGGGTGGCTGTGGCAATTTCTTAAA- 57957

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTG 1383
                  | |||   ||||     || ||| ||| |||||| |||||||||| | |||
Db      57956 -------GTAGACAACAATGAAGTTTCACTTTTCTTTTCATAAAAGATTTCTTTGTAGCA 57904

Qy       1384 TGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG-ATCAAT 1442
                | |||||| ||||||||||||||  |||||||||  ||||| |||| ||| || |||
Db      57903 CGAGATGCTGTTTGATAGCATTTTACCCACAGTAGAAACTCTTTCAAAACTGGAATGAAT 57844

Qy       1443 CCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGT 1502
                || |||||||| |||     || || |  |||||| || ||||| ||| ||||| ||||
Db      57843 CCACTCAAACCATGCTGCTGCTTTACCAATTAAGTTTATGTAATACTCTAAATCCTTTGT 57784

Qy       1503 TGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGA 1562
                ||| ||||||||||  |||||||  ||||||||||||||||||||||||||      |
Db      57783 TGTTATTTCAACAAGGTTCACAGAATCTTCACCAGGAGTAGATTCCATCTCAAGAAAAAA 57724

Qy       1563 GATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAG 1622
                |    |  |   |  |||||||||| || ||||||||  |||| ||||| |||||||
Db      57723 ATTTATTTGCTCATTTGTCCATAAGAAGCAACTCCTCATCCATTCACGTTTTCTCATGAG 57664

Qy       1623 ATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGC 1682
                |||||||||| | ||||  | ||| ||     || ||||| |||||| ||||||||||
Db      57663 ATTGCAGCAATTCCGTCACATTTTCGGG-----CTCCACTTCTAATTCTAGTTCTCTTGC 57609

Qy       1683 TGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATC 1742
                |||||||||||||||  ||  |||||||  | |||||| ||||||| ||| || |||||
Db      57608 TGTTTCTACCACATCTGCAGTGACTTCCTAAACTGAAGTGTTGAACCCCTCAAATTCATC 57549

Qy       1743 CATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC-C 1801
                || |  |||||||| |||||| |||| | ||  |||| || ||  |||||||||| |
Db      57548 TATAACTGTTGGAATCAACTTCTTTCAAACTGCTACTAATGTTGATCCTTTGACCTCCTC 57489

Qy       1802 ATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCA 1861
                ||     |  |||||||| ||| || ||||||  |||||||| || ||||||||
Db      57488 CTGCGAATCACAAATTCTTAATAGCATCTAGAATGGCCAATCCTTTTCAGAAGGTTTTTG 57429

Qy       1862 ATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTT 1921
                |||||||| | |||| |||     |||||| |  |||| || ||  ||| |||||||||
Db      57428 ATTTACTTTGCCCAGGTCCA-----CCAGAGCAATCACTGTCTATAGCAGCTATAGCCTT 57374

Qy       1922 ATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTT 1981
                |   |||||| |              |  |||||| ||||||| ||||| |||| |||
Db      57373 ACAAAATGTACT-------------GATTTGAAACTTGAAATCACTCCCTGATTCATGGG 57327

Qy       1982 CTGCAAA 1988
                || ||||
Db      57326 CTACAAA 57320


RESULT 43
ADC85434/c
ID   ADC85434 standard; DNA; 96589 BP.
XX
AC   ADC85434;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human Nr3c1 genomic sequence.
XX
KW   Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW   secreted; transmembrane; intracellular; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003045230-A2.
XX
PD   05-JUN-2003.
XX
PF   02-DEC-2002; 2002WO-US038582.
XX
PR   30-NOV-2001; 2001US-00997722.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-513603/48.
XX
PT   New recombinant nucleic acid comprising a nucleotide sequence of any of
PT   the carcinoma-associated (CA) genes, useful for screening for drug
PT   candidates for diagnosing or treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 220; 983pp; English.
XX
CC   The invention relates to a recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the fully defined carcinoma-
CC   associated (CA) genes from the 50 tables given in the specification. The
CC   CA proteins are secreted, transmembrane or intracellular proteins. The
CC   recombinant nucleic acids are useful for screening for drug candidates
CC   for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC   ADC85514 represent CA genes of the invention.
XX
SQ   Sequence 96589 BP; 28270 A; 17369 C; 18773 G; 31440 T; 0 U; 737 Other;

  Query Match            22.1%;  Score 441;  DB 10;  Length 96589;
  Best Local Similarity  71.4%;  Pred. No. 1.1e-65;
  Matches  733;  Conservative    0;  Mismatches  230;  Indels   64;  Gaps    9;

Qy        974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
                || |||||||||||||  |||||| |   |||| |||||| | |||||| | |
Db      58294 TAATACAGGCATACCTCTGAGATATTTCAGGTTCGGTTCTAGACCACAATA--------- 58244

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
                          ||||| ||||  ||| ||||  ||||| || | | |||||| |||||| |
Db      58243 ---------AAGTGAATATTGCAACAAAGCAAGTCAGACGAATTTTTTGGTTTCCCACTA 58193

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
                |||||||||||| |  |||||||| |||||||||| ||||||||||||||   || ||||
Db      58192 CATATAAAAGTTATATTTATACTATACTGTAGTCTATTAAGTGTGCAATAGCATTGTGTC 58133

Qy       1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
                |  |||||   ||   |||  || |   |    |||| || ||||||||||||||||||
Db      58132 TTTAAAAATGTAT---TTACCTTAATCTTAAAATATTGCTTAAAAATGCTAACAATCATT 58076

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT----------TTTCTTA 1263
                ||| | ||||||||||||||||||||||| || ||||| |||||          ||  |
Db      58075 TGAACTTTCAGTGAGTTGTAATCTTTTTCCTTGTGGAGGGTCTTGCTTCAGTGTTGATGC 58016

Qy       1264 TTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA 1323
                || |||||||| | |    |||     |   | ||||||||||||| |||||||||
Db      58015 CTGCTGACTGATCAGAGAGGTGGTTGCTGAAGGTGGGGTGGCTGTGGCAATTTCTTAAA- 57957

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTG 1383
                  | |||   ||||     || ||| ||| |||||| |||||||||| | |||
Db      57956 -------GTAGACAACAATGAAGTTTCACTTTTCTTTTCATAAAAGATTTCTTTGTAGCA 57904
```

```
Qy       1384 TGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG-ATCAAT 1442
Db      57903 CGAGATGCTGTTTGATAGCATTTTACCCACAGTAGAAACTCTTTCAAAACTGGAATGAAT 57844

Qy       1443 CCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGT 1502
Db      57843 CCACTCAAACCATGCTGCTGCTTTACCAATTAAGTTTATGTAATACTCTAAATCCTTTGT 57784

Qy       1503 TGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGA 1562
Db      57783 TGTTATTTCAACAAGGTTCACAGAATCTTCACCAGGAGTAGATTCCATCTCAAGAAAAAA 57724

Qy       1563 GATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAG 1622
Db      57723 ATTTATTTGCTCATTTGTCCATAAGAAGCAACTCCTCATCCATTCACGTTTTCTCATGAG 57664

Qy       1623 ATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGC 1682
Db      57663 ATTGCAGCAATTCCGTCACATTTTCGGG-----CTCCACTTCTAATTCTAGTTCTCTTGC 57609

Qy       1683 TGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATC 1742
Db      57608 TGTTTCTACCACATCTGCAGTGACTTCCTAAACTGAAGTGTTGAACCCCTCAAATTCATC 57549

Qy       1743 CATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC-C 1801
Db      57548 TATAACTGTTGGAATCAACTTCTTTCAAACTGCTACTAATGTTGATCCTTTGACCTCCTC 57489

Qy       1802 ATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCA 1861
Db      57488 CTGCGAATCACAAATTCTTAATAGCATCTAGAATGGCCAATCCTTTTCAGAAGGTTTTTG 57429

Qy       1862 ATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTT 1921
Db      57428 ATTTACTTTGCCCAGGTCCA-----CCAGAGCAATCACTGTCTATAGCAGCTATAGCCTT 57374

Qy       1922 ATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTT 1981
Db      57373 ACAAAATGTACT------------GATTTGAAACTTGAAATCACTCCCTGATTCATGGG 57327

Qy       1982 CTGCAAA 1988
Db      57326 CTACAAA 57320
```

RESULT 44
ADM74549/c
ID ADM74549 standard; DNA; 96589 BP.
XX
AC ADM74549;
XX
DT 01-JUL-2004 (first entry)
XX
DE Human carcinoma associated (CA) nucleic acid #109.
XX
KW Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW cytostatic.
XX
OS Homo sapiens.
XX
PN US2004072154-A1.
XX
PD 15-APR-2004.
XX
PF 30-NOV-2001; 2001US-00997722.
XX
PR 22-DEC-2000; 2000US-00747377.
PR 02-MAR-2001; 2001US-00798586.
XX
PA (MORR/) MORRIS D W.
PA (ENGE/) ENGELHARD E K.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2004-328562/30.
XX
PT New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
PT lymphoma.
XX
PS Claim 1; SEQ ID NO 220; 29pp; English.
XX
CC The invention relates to new recombinant nucleic acids. The invention
CC also relates to a host cell comprising a recombinant nucleic acid or
CC expression vector, an expression vector comprising a recombinant nucleic
CC acid, a recombinant protein, a method of screening for drug candidates, a
CC method of screening for a bioactive agent capable of binding to a
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC method of screening for a bioactive agent capable of modulating the
CC activity of a CAP, a method of evaluating the effect of a candidate
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC the activity of a CAP, a method of treating carcinomas, a method of
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC propensity to carcinoma. A method of evaluating the effect of a candidate
CC carcinoma drug comprises administering the drug to a patient, removing a
CC cell sample from the patient and determining alterations in the
CC expression or activation of a gene comprising the nucleotide sequence. A
CC method of diagnosing carcinoma comprises determining the expression of
CC one or more genes comprising the nucleic acid sequence in a first tissue
CC type of a first individual and comparing the expression of the gene from
CC a second normal tissue type from the first individual or a second
CC unaffected individual, where a difference in the expression indicates
CC that the first individual has carcinoma. A method of inhibiting the
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC carcinomas comprises administering to a patient an inhibitor of CAP.
CC Neutralising the effect of a CAP comprises contacting an agent specific
CC for the CAP. The polypeptide specifically binds to the protein encoded by
CC the nucleic acid. It comprises an antibody that specifically binds to the
CC protein encoded by the nucleic acid. The nucleic acids are useful for
CC preparing a composition for diagnosing or treating carcinoma e.g.,
CC leukaemia or lymphoma. This sequence represents a human carcinoma
CC associated (CA) nucleic acid of the invention. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 96589 BP; 28270 A; 17369 C; 18773 G; 31440 T; 0 U; 737 Other;

Query Match 22.1%; Score 441; DB 12; Length 96589;
Best Local Similarity 71.4%; Pred. No. 1.1e-65;
Matches 733; Conservative 0; Mismatches 230; Indels 64; Gaps 9;

```
Qy        974 TAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAA 1033
Db      58294 TAATACAGGCATACCTCTGAGATATTTCAGGTTCGGTTCTAGACCACAATA--------- 58244

Qy       1034 ATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTG 1093
Db      58243 ---------AAGTGAATATTGCAACAAAGCAAGTCAGACGAATTTTTTGGTTTCCCACTA 58193

Qy       1094 CATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTC 1153
Db      58192 CATATAAAAGTTATATTTATACTATACTGTAGTCTATTAAGTGTGCAATAGCATTGTGTC 58133

Qy       1154 TAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
Db      58132 TTTAAAAATGTAT---TTACCTTAATCTTAAAATATTGCTTAAAAATGCTAACAATCATT 58076

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCT----------TTTCTTA 1263
Db      58075 TGAACTTTCAGTGAGTTGTAATCTTTTTCCTTGTGGAGGGTCTTGCTTCAGTGTTGATGC 58016

Qy       1264 TTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA 1323
```

```
Db      58015 CTGCTGACTGATCAGAGAGGTGGTTGCTGAAGGTGGGGTGGCTGTGGCAATTTCTTAAA- 57957

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTG 1383

Db      57956 -------GTAGACAACAATGAAGTTTCACTTTTCTTTTCATAAAAGATTTCTTTGTAGCA 57904

Qy       1384 TGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG-ATCAAT 1442

Db      57903 CGAGATGCTGTTTGATAGCATTTTACCCACAGTAGAAACTCTTTCAAAACTGGAATGAAT 57844

Qy       1443 CCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGT 1502

Db      57843 CCACTCAAACCATGCTGCTGCTTTACCAATTAAGTTTATGTAATACTCTAAATCCTTTGT 57784

Qy       1503 TGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGA 1562

Db      57783 TGTTATTTCAACAAGGTTCACAGAATCTTCACCAGGAGTAGATTCCATCTCAAGAAAAAA 57724

Qy       1563 GATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAG 1622

Db      57723 ATTTATTTGCTCATTTGTCCATAAGAAGCAACTCCTCATCCATTCACGTTTTCTCATGAG 57664

Qy       1623 ATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGC 1682

Db      57663 ATTGCAGCAATTCCGTCACATTTTCGGG-----CTCCACTTCTAATTCTAGTTCTCTTGC 57609

Qy       1683 TGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATC 1742

Db      57608 TGTTTCTACCACATCTGCAGTGACTTCCTAAACTGAAGTGTTGAACCCCTCAAATTCATC 57549

Qy       1743 CATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC-C 1801

Db      57548 TATAACTGTTGGAATCAACTTCTTTCAAACTGCTACTAATGTTGATCCTTTGACCTCCTC 57489

Qy       1802 ATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCA 1861

Db      57488 CTGCGAATCACAAATTCTTAATAGCATCTAGAATGGCCAATCCTTTTCAGAAGGTTTTTG 57429

Qy       1862 ATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTT 1921

Db      57428 ATTTACTTTGCCCAGGTCCA-----CCAGAGCAATCACTGTCTATAGCAGCTATAGCCTT 57374

Qy       1922 ATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTT 1981

Db      57373 ACAAAATGTACT------------GATTTGAAACTTGAAATCACTCCCTGATTCATGGG 57327

Qy       1982 CTGCAAA 1988

Db      57326 CTACAAA 57320
```

RESULT 45
ADF13116
ID ADF13116 standard; DNA; 128779 BP.
XX
AC ADF13116;
XX
DT 12-FEB-2004 (first entry)
XX
DE Hypermethylation site in human breast cancer CpG island locus HBC-45.
XX
KW DNA methylation; CpG dinucleotide rich region;
KW differential methylation hybridisation; DMH; CpG island; screening array;
KW breast cancer; prostate cancer; colon cancer; lung cancer; liver cancer;
KW ovarian cancer; human; HBC; hypermethylation in breast cancer; ds.
XX
OS Homo sapiens.
XX
PN US2003129602-A1.
XX
PD 10-JUL-2003.
XX
PF 21-FEB-2002; 2002US-00081327.
XX
PR 05-FEB-1999; 99US-0118760P.
PR 18-FEB-1999; 99US-0120592P.
PR 04-FEB-2000; 2000US-00497855.
XX
PA (HUAN/) HUANG T H.
XX
PI Huang TH;
XX
DR WPI; 2004-031298/03.
XX
PT Differential methylation hybridization assay for detecting
PT hypermethylation of CpG dinucleotide rich regions in genomic DNA is
PT useful to diagnose and monitor breast, prostate, colon, lung, liver and
PT ovarian cancer.
XX
PS Claim 17; SEQ ID NO 38; 32pp; English.
XX
CC The present invention relates to a method of detecting DNA methylation of
CC a CpG dinucleotide rich region of a nucleic acid. The method comprises
CC hybridisation by differential methylation hybridisation (DMH) in which
CC the nucleic acid is digested into fragments in which CpG islands are
CC preserved, attaching the fragments to linker primers, digesting the
CC linker primer products to digest unmethylated CpG sequences, amplifying
CC and labelling the remaining linker primer products and detecting labelled
CC amplicons. Also disclosed is a screening array comprising nucleic acid
CC fragments fixed to a solid support where each fragment is a CpG
CC dinucleotide rich fragment comprising at least 200 nucleotides of which
CC at least 50% are guanine and cytosine. Nucleic acid probes prepared from
CC a cell sample are used to screen CpG dinucleotide rich fragments fixed
CC onto a screening array. The invention is useful to diagnose and monitor
CC prognosis of a disease associated with aberrant DNA methylation,
CC particularly breast, prostate, colon, lung, liver or ovarian cancer,
CC especially breast cancer. The present sequence represents the site of
CC hypermethylation in human breast cancer at a CpG island locus (HBC,
CC hypermethylation in breast cancer).
XX
SQ Sequence 128779 BP; 38014 A; 25995 C; 25782 G; 38988 T; 0 U; 0 Other;

Query Match 22.0%; Score 439.8; DB 12; Length 128779;
Best Local Similarity 72.7%; Pred. No. 1.7e-65;
Matches 747; Conservative 0; Mismatches 232; Indels 48; Gaps 12;

```
Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036

Db      42749 TACAGGCATACATTGGACATATTGCAGATTTGGTTCTAGACCACTGCAAAAAAGCAAATA 42808

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096

Db      42809 TTG---------------CAATAAAGTGAGTTACACAATTTTTTTGCCTTTCCAGTGCAT 42853

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156

Db      42854 ATAAAAGTTATGTTTATACTACATGACAGTCTATTAAGTGTGCAGTAGCAT--GTCTAAA 42911

Qy       1157 AAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGA 1216

Db      42912 AAATGCATATACCTTAATTAAAAGATACTTTATTGCTAAAAAATGC-AGGACTCTTCACT 42970

Qy       1217 GCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGA-T 1275

Db      42971 GAGTCATAATCTTTGAGGCTGCAGAGTCTTACCTTAATGTCTGCTGACTCATGGCTGATT 43030

Qy       1276 CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGT 1330

Db      43031 AGGGTGATAACTGCTGAAGGCTCAGGTGGCTGTGGCAATTTCTTAAAACATGACAACACT 43090

Qy       1331 GAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATG 1390

Db      43091 GAAGTTTGCTGCATCAGTTGATTCTTCCTTTCATGAACTATTTCTCTGTGGTATGTGATG 43150

Qy       1391 CTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAA 1450
```

```
Db      43151 CTGTTTAATAGCATTTTACCCACAGAA-----CTCTCAAAGTTGGAGTCAATCCTCTCAA 43205

Qy       1451 ACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTT 1510
              || ||||       || |  |   ||||| || | | ||||| ||| | ||| ||| ||||
Db      43206 ACTCTGCCACTGCTTTATCAGTCAAGTTTATGTGACATTCTAAATTCTTTGCTGTTATTT 43265

Qy       1511 CAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAAT 1570
              |||||| ||||||||  ||||| |||||||||||||| ||||  |        |  |    |
Db      43266 CAACAA-TTTCACAGCATCTTCCCCAGGAGTAGATTTCATCCTAAGAAACCACTT----T 43320

Qy       1571 CTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGC 1630
              |||| |||||||||||||| || | ||||||  ||||| ||||||||| ||| ||||||
Db      43321 CTTTACTCATCCATAAGAAGCAACTTCTCATCCATTCAAATTTTATCATAAGAATGCAGC 43380

Qy       1631 AATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTA 1690
              ||| | ||||  ||||||||     || ||||| |||||| |||||||||||| |||| |
Db      43381 AATTCTGTCACATCTTCAGG-----CTACACTTCTAATTCTAGTTCTCTTGCTATTTCCA 43435

Qy       1691 CCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGG 1750
              | | ||||   |||  |||||||| |||||||||||||| |  ||||||||||||||
Db      43436 CTATATCTACAGTTATTTCCTCCACTGAAGTCTTGAACCTGTGAAAGTCATCCATGAGGA 43495

Qy       1751 TTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAAT 1807
              | |||||| |||||||||||| ||||||| || || |||||||||   ||||||||||||
Db      43496 TGGGAATCAACTTCTTCCAAACTCCTGTTCATGTTGATATTTTGATCTCCTCCCATGAAT 43555

Qy       1808 CATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTAC 1867
              ||  |||||||||||| ||| || |||| |||||||||| |||| |||||||||||| |||
Db      43556 CACAAATGTTCTTAATTGCATCTAGAAT-GTGAATCCTCTCCAGAAGGTTTTCAATGTAC 43614

Qy       1868 TTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAA 1927
              || |  |||||||||     |||||||  |||| || ||| ||| |||| |||||   ||
Db      43615 TTTGCTCAGATCCAT-----CAGAGGAATCACTATCTATGGCAGCTATACCCTTACAAAA 43669

Qy       1928 TGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAA 1987
              ||||||||| |||||||| |||||||||||||||||||||||||  |   |||
Db      43670 TGTATTTCTTAAATAATAAGACTTGAAAGTTGAAATTACTCCTTGATCTGTGGGCTGTGG 43729

Qy       1988 AATAGAT 1994
              ||| | |
Db      43730 AATGGTT 43736
```

Search completed: January 22, 2006, 04:08:38
Job time : 831.678 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:39:03 ; Search time 245.102 Seconds
(without alignments)
14504.671 Million cell updates/sec

Title: US-09-728-552A-3_COPY_1_2000
Perfect score: 2000
Sequence: 1 gaattctcctgcctcagcct.....tctgcaaaatagatgttgtg 2000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2000	100.0	80595	3	US-09-078-294-3	Sequence 3, Appli
	2	1611.4	80.6	80246	3	US-09-078-294-4	Sequence 4, Appli
c	3	636.8	31.8	2418	3	US-09-573-080A-148	Sequence 148, App
	4	483.8	24.2	26416	3	US-09-949-016-14266	Sequence 14266, A
c	5	480	24.0	119762	3	US-09-949-016-17313	Sequence 17313, A
c	6	472.4	23.6	237510	3	US-09-949-016-14273	Sequence 14273, A
c	7	468.2	23.4	360470	3	US-09-949-016-13173	Sequence 13173, A
	8	465.8	23.3	235064	3	US-09-949-016-15390	Sequence 15390, A
	9	465.6	23.3	164061	3	US-09-949-016-17422	Sequence 17422, A
c	10	462.6	23.1	194714	3	US-09-949-016-11869	Sequence 11869, A
c	11	462.6	23.1	196714	3	US-09-949-016-15474	Sequence 15474, A
	12	461.2	23.1	340380	3	US-09-949-016-14179	Sequence 14179, A
	13	454.2	22.7	260247	3	US-09-949-016-13358	Sequence 13358, A
c	14	453.8	22.7	59948	3	US-09-949-016-15594	Sequence 15594, A
c	15	453.4	22.7	135687	3	US-09-949-002-805	Sequence 805, App
	16	453	22.7	145812	3	US-09-949-016-15698	Sequence 15698, A
	17	446	22.3	92681	3	US-09-949-016-14772	Sequence 14772, A
c	18	444.2	22.2	264206	3	US-09-949-016-12731	Sequence 12731, A
c	19	444.2	22.2	264304	3	US-09-949-016-13249	Sequence 13249, A
c	20	443	22.1	177797	3	US-09-949-016-14125	Sequence 14125, A
c	21	443	22.1	227979	3	US-09-949-016-11842	Sequence 11842, A
	22	439.8	22.0	128779	3	US-09-497-855A-38	Sequence 38, Appl
	23	438.4	21.9	199471	3	US-09-949-016-14083	Sequence 14083, A
	24	437.4	21.9	301828	3	US-09-949-016-13969	Sequence 13969, A
c	25	437.4	21.9	385136	3	US-09-949-016-16073	Sequence 16073, A
c	26	435.8	21.8	96202	3	US-09-949-016-16433	Sequence 16433, A
	27	434	21.7	103750	3	US-09-949-016-13319	Sequence 13319, A
	28	433.6	21.7	113701	3	US-09-949-016-13214	Sequence 13214, A
c	29	433	21.6	95750	3	US-09-949-016-11926	Sequence 11926, A
c	30	430.8	21.5	112239	3	US-09-949-016-13144	Sequence 13144, A
c	31	429	21.4	343352	3	US-09-949-016-13498	Sequence 13498, A
	32	428.4	21.4	321022	3	US-09-949-016-11852	Sequence 11852, A
	33	428.4	21.4	321022	3	US-09-949-016-14166	Sequence 14166, A
	34	427.8	21.4	61461	3	US-09-949-016-16419	Sequence 16419, A
c	35	426	21.3	818128	3	US-09-949-016-14546	Sequence 14546, A
c	36	426	21.3	818128	3	US-09-949-016-14547	Sequence 14547, A
c	37	426	21.3	818128	3	US-09-949-016-14548	Sequence 14548, A
c	38	426	21.3	818128	3	US-09-949-016-14549	Sequence 14549, A
c	39	426	21.3	818128	3	US-09-949-016-14550	Sequence 14550, A
c	40	426	21.3	818128	3	US-09-949-016-14551	Sequence 14551, A
c	41	426	21.3	818128	3	US-09-949-016-14552	Sequence 14552, A
c	42	426	21.3	818128	3	US-09-949-016-14553	Sequence 14553, A
c	43	426	21.3	818128	3	US-09-949-016-14554	Sequence 14554, A
c	44	426	21.3	818128	3	US-09-949-016-14555	Sequence 14555, A
c	45	426	21.3	818128	3	US-09-949-016-14556	Sequence 14556, A
c	46	426	21.3	818128	3	US-09-949-016-14557	Sequence 14557, A
c	47	426	21.3	818128	3	US-09-949-016-14558	Sequence 14558, A
c	48	426	21.3	818128	3	US-09-949-016-14559	Sequence 14559, A
c	49	426	21.3	818128	3	US-09-949-016-14560	Sequence 14560, A
c	50	426	21.3	818128	3	US-09-949-016-14561	Sequence 14561, A
c	51	426	21.3	818128	3	US-09-949-016-14562	Sequence 14562, A
c	52	426	21.3	818128	3	US-09-949-016-14564	Sequence 14564, A
c	53	426	21.3	818128	3	US-09-949-016-14565	Sequence 14565, A
c	54	426	21.3	818128	3	US-09-949-016-14566	Sequence 14566, A
c	55	426	21.3	818128	3	US-09-949-016-14567	Sequence 14567, A
	56	424.2	21.2	88490	3	US-09-949-016-12758	Sequence 12758, A
	57	424.2	21.2	88736	3	US-09-949-016-14222	Sequence 14222, A
	58	422.6	21.1	175236	3	US-09-949-016-14353	Sequence 14353, A
	59	421.8	21.1	450395	3	US-09-949-016-15473	Sequence 15473, A
	60	420	21.0	142783	3	US-09-949-016-15127	Sequence 15127, A
	61	418.8	20.9	25814	3	US-09-949-016-16927	Sequence 16927, A
	62	418	20.9	27702	3	US-09-949-016-11795	Sequence 11795, A
c	63	416	20.8	112132	3	US-09-741-150-3	Sequence 3, Appli
c	64	416	20.8	112132	3	US-10-160-187-3	Sequence 3, Appli
c	65	413.4	20.7	63860	3	US-09-949-016-15825	Sequence 15825, A
	66	413.4	20.7	264206	3	US-09-949-016-12731	Sequence 12731, A
	67	413.4	20.7	264304	3	US-09-949-016-13249	Sequence 13249, A
	68	411.6	20.6	96202	3	US-09-949-016-16433	Sequence 16433, A
c	69	411.2	20.6	198632	3	US-09-949-016-12781	Sequence 12781, A
c	70	411.2	20.6	198637	3	US-09-949-016-17393	Sequence 17393, A
c	71	411	20.5	88758	3	US-09-949-016-13502	Sequence 13502, A
c	72	410.6	20.5	285478	3	US-09-949-016-13362	Sequence 13362, A
	73	408	20.4	300402	3	US-09-949-016-13632	Sequence 13632, A
c	74	406.2	20.3	76269	3	US-09-949-016-14603	Sequence 14603, A
c	75	406.2	20.3	99304	3	US-09-949-016-15440	Sequence 15440, A
c	76	401.6	20.1	160018	3	US-09-949-016-12617	Sequence 12617, A
c	77	401.6	20.1	160018	3	US-09-949-016-15994	Sequence 15994, A
	78	401	20.1	89268	3	US-09-949-016-12025	Sequence 12025, A
	79	401	20.1	90544	3	US-09-949-016-13302	Sequence 13302, A
c	80	400	20.0	83665	3	US-09-949-016-16995	Sequence 16995, A
	81	397.8	19.9	36652	3	US-09-949-016-14683	Sequence 14683, A
	82	397.4	19.9	66428	3	US-09-949-016-12917	Sequence 12917, A
	83	397.4	19.9	107679	3	US-09-949-016-16409	Sequence 16409, A
	84	396.6	19.8	237863	3	US-09-949-016-13404	Sequence 13404, A
	85	395.8	19.8	5848	3	US-09-949-016-16556	Sequence 16556, A
	86	395.8	19.8	6623	3	US-09-949-016-12157	Sequence 12157, A
	87	395.8	19.8	109025	3	US-09-949-016-12609	Sequence 12609, A
	88	395.8	19.8	109025	3	US-09-949-016-17567	Sequence 17567, A
c	89	395.8	19.8	113701	3	US-09-949-016-13214	Sequence 13214, A
c	90	395.2	19.8	94987	3	US-09-949-016-12510	Sequence 12510, A
	91	395.2	19.8	228896	3	US-09-949-016-17127	Sequence 17127, A
c	92	394.6	19.7	260286	3	US-09-949-016-17037	Sequence 17037, A
c	93	394.6	19.7	260293	3	US-09-949-016-12106	Sequence 12106, A
	94	394.6	19.7	387902	3	US-09-949-016-14543	Sequence 14543, A
	95	394.6	19.7	421883	3	US-09-949-016-12557	Sequence 12557, A
c	96	394	19.7	124110	3	US-09-949-016-13353	Sequence 13353, A
	97	392.2	19.6	360470	3	US-09-949-016-13173	Sequence 13173, A

c 98	391	19.6	153642	3	US-09-949-016-12174	Sequence 12174, A
c 99	391	19.6	153643	3	US-09-949-016-15635	Sequence 15635, A
100	389.6	19.5	455726	3	US-09-949-016-14157	Sequence 14157, A

ALIGNMENTS

RESULT 1
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match 100.0%; Score 2000; DB 3; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCA 60

Qy         61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAA 120

Qy        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTG 180

Qy        181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAG 240

Qy        241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTTATTATTAGGGGCACATACATTTAAAATTGTTATGTCTTATTGATAGATTGATCT 300

Qy        301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTTCTAAATATTTTTCTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTTCTAAATATTTTTCTG 360

Qy        361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTTTATGGTATA 420

Qy        421 TATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATACTTAAAGTGG 480

Qy        481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGATAATCTCAAC 540

Qy        541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATGGTTTGATTT 600

Qy        601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTCTTTTTTCAT 660

Qy        661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATCTATTGGCTT 720

Qy        721 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAATATACATCT 780

Qy        781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAATGTAGAAAC 840

Qy        841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATGTCTATAATACATT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATGTCTATAATACATT 900

Qy        901 AGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCAAACATTTTCT 960

Qy        961 GTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCAC 1020

Qy       1021 CACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTT 1080

Qy       1081 TGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCA 1140

Qy       1141 ATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAAT 1200

Qy       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC 1260

Qy       1261 TTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTA 1320

Qy       1321 AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCA 1440

Qy       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500

Qy       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560

Qy       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620

Qy       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680

Qy       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740

Qy       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCC 1800

Qy       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTC 1860

Qy       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCT 1920

Qy       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTT 1980

Qy       1981 TCTGCAAAATAGATGTTGTG 2000
             ||||||||||||||||||||
Db       1981 TCTGCAAAATAGATGTTGTG 2000


RESULT 2
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

  Query Match             80.6%;  Score 1611.4;  DB 3;  Length 80246;
  Best Local Similarity   97.4%;  Pred. No. 0;
  Matches 1666;  Conservative    2;  Mismatches   40;  Indels    3;  Gaps    3;

Qy        292 GATTGATCTGTCATTATGAATGTCTGTATTCATTCCCTGATAGTATTTCTTTTTCTAAAT 351
              | ||||| |||   || | | || || | ||| |||| ||  | |||| |||||  |||
Db          2 GGTTGATTTGTNNATAAGGAAGTTTGGAATCAATCCCGGAAGGAATTTTTTTTTTAAAAA 61

Qy        352 ATTTTTCTGAATGTGTCTGCTATTAACATAGCCACTCTGGCTTTTTAAAATTAGTATTTT 411
              ||||||  ||| | || || ||:||| | |:||| | ||| ||||||||| ||| | |||
Db         62 ATTTTTTGGAAGG-GTTTGGTAWTAAAAAARCCAATTTGGGTTTTTAAAAATAGGAATTT 120

Qy        412 TATGGTATATATTTTTCCTTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATAC 471
              ||||| | | | |||||| ||||||||||||||||||||||||||||||||||||||||||
Db        121 TATGGGAAAAAATTTTCCCTTTTTTTTTTTTTTAAGTTTTAGATGTTATGTTTCCTTATAC 180

Qy        472 TTAAAGTGGGTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGAT 531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTAAAGTGGGTGTCTTATAGGCAGCATATATCTGGGTCTTGATGTATTATTTAATCTGAT 240

Qy        532 AATCTCAACCTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATG 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AATCTCAACCTTTTTGTTGGAGTGTTTAGGCCATTTACATTTAGTGTAATTATAGACATG 300

Qy        592 GTTTGATTTGCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTC 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTTTGATTTGCTATACCATCTTTTCATTTGTTTTATATGTGAGCCATCTTTTCATTGTTC 360

Qy        652 TTTTTTCATCTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATC 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        361 TTTTTTCATCTTTGACCATTTTCTTTAGTACTGAATACTTTTTTTGTATTTCATTATATC 420

Qy        712 TATTGGCTTTTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAA 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TATTGGCTTTTTAGTTATACCTCTTAAAATTTTTTTTTCTGTTTTATGTAGGATTTATAA 480

Qy        772 TATACATCTTTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAA 831
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TATACATCTTTAACTTATCACAGATTACCTTCAAATAGTATTTTACCAGCTCAAGTGTAA 540

Qy        832 TGTAGAAACCTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTA-TGTCT 890
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        541 TGTAGAAACCTTACAAGAGTATATTTTCATTTCTGTCTCCTAATTTTTATGCTATTGTCT 600

Qy        891 ATAATACATTAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCA 950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATAATACATTAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTGTTGGCTGGGGTCAGCA 660

Qy        951 AACATTTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGT 1010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACATTTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGT 720

Qy       1011 TCCATACCACCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCAC 1070
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCCATACCACCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCAC 780

Qy       1071 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 1130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGT 840

Qy       1131 TAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATT 1190
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATT 900

Qy       1191 ACTAAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGA 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTAAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGA 960

Qy       1251 AGGTCTTTTCTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGG 1310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AGGTCTTTTCTTATTGATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGG 1020

Qy       1311 CAGTTTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGA 1370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGTTTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGA 1080

Qy       1371 TTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAA 1430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAA 1140

Qy       1431 AATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATT 1489
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATTGGAGTCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATT 1200

Qy       1490 CTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCA 1549
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCA 1260

Qy       1550 TCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAA 1609
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAA 1320

Qy       1610 GTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATT 1669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATT 1380

Qy       1670 CCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACC 1729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACC 1440

Qy       1730 TCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATA 1789
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1441 TCTCCAAGTCATCCATGAGGGCTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATA 1500
```

```
Qy       1790 TTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCC 1849
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TTTTGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCC 1560

Qy       1850 AAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCC 1909
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCC 1620

Qy       1910 AGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCC 1969
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCC 1680

Qy       1970 TTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
             ||||||||||||||||||||||||||||||
Db       1681 TTGATCCATTTTCTGCAAAATAGATGTTGTG 1711


RESULT 3
US-09-573-080A-148/c
; Sequence 148, Application US/09573080A
; Patent No. 6828097
; GENERAL INFORMATION:
;  APPLICANT:  JOAN, KNOLL
;  APPLICANT:  ROGAN, PETER
;  TITLE OF INVENTION:  SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATI
;  FILE REFERENCE:  30307
;  CURRENT APPLICATION NUMBER: US/09/573,080A
;  CURRENT FILING DATE:  2000-05-16
;  NUMBER OF SEQ ID NOS:  479
;  SOFTWARE:  PatentIn version 3.0
; SEQ ID NO 148
;  LENGTH: 2418
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: repeat_region
;  LOCATION: (1)..(2418)
;  OTHER INFORMATION: trigger1
;  PUBLICATION INFORMATION:
;  PUBLICATION INFORMATION:
;  AUTHORS: Jurka, J; Walichiewicz, J; Milosavljevic, A
;  TITLE: Prototypic sequences for human repetitive DNA
;  JOURNAL: Journal of Molecular Evolution
;  VOLUME: 35
;  ISSUE: 4
;  PAGES: 286-291
;  DATE: 1992-10-__
;  DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
;  DATABASE ENTRY DATE:  ____-__-__
;  DATABASE ENTRY DATE: 1996-01-26
US-09-573-080A-148

  Query Match             31.8%;  Score 636.8;  DB 3;  Length 2418;
  Best Local Similarity   82.1%;  Pred. No. 1e-121;
  Matches  859;  Conservative    1;  Mismatches  133;  Indels   53;  Gaps    9;

Qy        979 CAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATG 1038
              |||||||||||| ||||||| || ||||| ||||||| |||||| |||||
Db       2418 CAGGCATACCTCGGAGATATTGCGGGTTCGGTTCCAGACCACCGCAATA---------- 2370

Qy       1039 CAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATAT 1098
                 ||| | ||||| |||||||| ||||||||| | | |||||| ||||||||||||||
Db       2369 ----AAGCGAATATCGCAATAAAGCGAGTCACACGAATTTTTTGGTTTCCCAGTGCATAT 2314

Qy       1099 AAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAA 1158
              ||||||| || |||  ||| |||||||||| ||||||||||||||  ||||||||||||
Db       2313 AAAAGTTATGTTTACGCTATACTGTAGTCTATTAAGTGTGCAATAGCATTATGTCTAAAA 2254

Qy       1159 AA----ACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTT 1214
              ||    ||||||||||||||| ||||| ||||||| ||||||||||||||| ||||| |
Db       2253 AACAATGTACATACCTTAATTTAAAAATACTTTATTGCTAAAAAATGCTAACGATCATCT 2194

Qy       1215 GAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGA 1274
              |||| ||||| |||| |||||||||||||||||||| ||||||  |||  || |||  |
Db       2193 GAGCCTTCAGCGAGTCGTAATCTTTTTGCTGGTGGAGGGTCTTGCCTTGATGTTGATGGC 2134

Qy       1275 T-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA- 1322
              |           ||| |   | |||||||| || ||||||||||||| |||||||||
Db       2133 TGCTGACTGATCAGGGTGGTGGTTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCTTAAAA 2074

Qy       1323 ----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTC 1378
                  |||||| ||||| ||||   |||| ||||||||||||||||||||||||||||||||
Db       2073 TAAGACAACAATGAAGTTTGCCGCATCAATTGACTCTTCCTTTCATGAAAGATTTCTCTG 2014

Qy       1379 TAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA- 1437
              |||  ||||||||| ||||| |||||||||  |||||||||||||||| |||||||||
Db       2013 TAGCATGTGATGCTGTTTGACAGCATTTTACCCACAGTAGAACTTCTTTCAAAATTGGAG 1954

Qy       1438 TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCC 1497
              |||||||||||||||||||      || |  | ||||||| || |||||||| |||||
Db       1953 TCAATCCTCTCAAACCCTGCCGCTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCC 1894

Qy       1498 ATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTT 1557
              |||||||||||||||||| |||||||  ||||||||||||||||||||||||||||||
Db       1893 TTTGTTGTCATTTCAACAATGTTCACAGCATCTTCACCAGGAGTAGATTCCATCTCAAGA 1834

Qy       1558 CCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATC 1617
              |  |  ||||||||||||||||||||| || |||||||| ||||||||||||||||
Db       1833 AACCACTT----TCTTTGCTCATCCATAAGAAGCAACTCCTCATCCGTTCAAGTTTTATC 1778

Qy       1618 ATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCT 1677
              |||||||||||||||| ||||||  ||||||||      || ||||| |||||| ||||||
Db       1777 ATGAGATTGCAGCAATTCAGTCACATCTTCAGG-----CTCCACTTCTAATTCTAGTTCT 1723

Qy       1678 CTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAG 1737
              |||||| |||| |||||||||| ||| |||||||| |||||||||:|||| ||| |||
Db       1722 CTTGCTATTTCCACCACATCTGCAGTTACTTCCTCCACTGAAGTCTTRAACCCCTCAAAG 1663

Qy       1738 TCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA-- 1795
              |||||||||||||||||||| |||||||||| |||||||||  || |||||||||
Db       1662 TCATCCATGAGGGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGATATTTTGACC 1603

Qy       1796 -CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAG 1854
              |||||||||||||| |||||||||||||||||| || |||||||||||||||||||| |||
Db       1602 TCCTCCCATGAATCACGAATGTTCTTAATGGCATCTAGAATGGTGAATCCTTTCCAGAAG 1543

Qy       1855 GTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTA 1914
              |||||||||||||| | ||||||||||      |||||||  |||| || ||| ||| ||
Db       1542 GTTTTCAATTTACTTTGCCCAGATCCAT-----CAGAGGAATCACTATCTATGGCAGCTA 1488

Qy       1915 TAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGAT 1974
              ||||||||| |||||||||||| ||||||||| ||||||||| ||||||||||||||||
Db       1487 TAGCCTTATAAAATGTATTTCTTAAATAATAAGACTTGAAAGTCGAAATTACTCCTTGAT 1428

Qy       1975 CCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||   ||||| ||| |||||||||
Db       1427 CCATGGGCTGCAGAATGGATGTTGTG 1402


RESULT 4
US-09-949-016-14266
; Sequence 14266, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
```

```
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14266
;   LENGTH: 26416
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-14266

  Query Match            24.2%;  Score 483.8;  DB 3;  Length 26416;
  Best Local Similarity  74.6%;  Pred. No. 4.4e-90;
  Matches  769;  Conservative    0;  Mismatches  192;  Indels   70;  Gaps   10;

Qy        978 ACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATAT 1037
              |||| || ||  |||| ||| |   ||||  |||||| | ||| |||||||  |||
Db      13527 ACAGTCACACACTGGACATATTTCAGGTTCAGTTCCAGAACACTACAATAAAGCAA---- 13582

Qy       1038 GCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATA 1097
                 ||||||||||||||  |||||||  | | | |||| ||| |||||||||
Db      13583 -----------ATATCACAATAAAGCAAGTCACATGAATTTGTTGGTTTCTCAGTGCATA 13631

Qy       1098 TAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAA 1157
              ||||||||||  ||||                  | | | |||||||  |||||||| ||
Db      13632 TAAAAGTTTTGTCTATA-----------------AGGCGAGCAATAGCATTATGTCTGAA 13674

Qy       1158 AAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAG 1217
              |||  || ||| ||||||  ||||| ||||||| || ||||| | ||||| |||| ||||
Db      13675 AAAGTACTTACTTTAATTAAAAAATACTTTATTGCTTAAAAAGGTTAACAGTCATCTGAG 13734

Qy       1218 CATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-- 1275
              | || ||  |||||||||||||||||||||||| ||||||  || | ||   | || |
Db      13735 CCTTTAGCAAGTTGTAATCTTTTTGCTGGTGGATGGTCTTGCCTCAGTGTGAATTGCTGC 13794

Qy       1276 ---------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT-----AA 1321
                      |||  |    | ||||||| || ||||||||||||||| ||||||      ||
Db      13795 TGACTGATCAGGGTGGTTGTTGCTGAAGGTTGGGGTGGCTGTGGCAATTTCTTCAGAGAA 13854

Qy       1322 AACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAG 1381
              |||||| |||||  |||| | ||   ||||||| ||||||||  |||||||| ||| | |
Db      13855 GACAACAATGAAGTGTGCCACATTGATTGACTCCTCCTTTCACAAAAGATTTATCTTTGG 13914

Qy       1382 TGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCA 1440
              |||||||||  |||||||||||  ||    |||||||| |||||||  |||||||||  |||
Db      13915 CATGTGATGCTGTTTGATAGCATGTT----ACAGTAGAGCTTCTTTCAAAATTGGAGTCA 13970

Qy       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500
              ||| ||||||  | |||       || |  | |||| || || ||  ||||||||||  ||
Db      13971 GTCCCCTCAAATCTTGCCAAAGCTTTATTAACTAATTTTATGTATAATTCTGAATCCTTT 14030

Qy       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560
              |||||||| |||||||| |||||||  ||||||||||||||||| || ||  |  |
Db      14031 GTTGTCATCTCAACAATGTTCACAGCATCTTCACCAGGAGTAGATGCCTTCCCAAGTAAC 14090

Qy       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
              |  |    |||||| |||||||||||||| || ||||||||   ||||||||||||||
Db      14091 CACTT----TCTTTGTTCATCCATAAGAAGCAACTCCTCATCCACTCAAGTTTTATCATG 14146

Qy       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680
              | |||||||| | ||||||  ||||||||     || ||||| |||||| |||||||||
Db      14147 AAATTGCAGCATTTCAGTCACATCTTCAGG-----CTCCACTTCTAATTCTAGTTCTCTT 14201

Qy       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740
              ||| |||| ||||||||||  ||  |||||||| | |||||||||| | ||| ||||||
Db      14202 GCTATTTCCACCACATCTGCAGTGACTTCCTCCACTAAAGTCTTGAATCCCTCAAAGTCA 14261

Qy       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CC 1797
              |||||||||  |||||||  |||||||  ||| || ||||||||||   |||||||   ||
Db      14262 TCCATGAGGACTGGAATCACCTTCTTCTAAACTCTTGTTAATATTGGCATTTTGACTTCC 14321

Qy       1798 TCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTT 1857
              | |||| ||||| |||||||||||||||||| || ||| | ||||||||||||| ||||||
Db      14322 TTCCATAAATCACGAATGTTCTTAATGGCATCTAGAACGATGAATCCTTTCCAGAAGGTT 14381

Qy       1858 TTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAG 1917
              ||||||| |||| | ||||||||||     ||||| |  |||| |  ||| ||| |||||
Db      14382 TTCAATTGACTTTGCCCAGATCCAT-----CAGAGTAATCACTATTTATGGCAGCTATAG 14436

Qy       1918 CCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCA 1977
              |||||| |||||||||||| ||||||| | |||||||||   |||||||| ||||||||
Db      14437 TCTTATGAAATGTATTTCTTAAATAATATGACTTGAAAGTCAGAATTACTCTTTGATCCA 14496

Qy       1978 TTTTCTGCAAA 1988
              |   ||| |||
Db      14497 TGGGCTGTAAA 14507


RESULT 5
US-09-949-016-17313/c
; Sequence 17313, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17313
;   LENGTH: 119762
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17313

  Query Match            24.0%;  Score 480;  DB 3;  Length 119762;
  Best Local Similarity  73.9%;  Pred. No. 3.7e-89;
  Matches  776;  Conservative    0;  Mismatches  220;  Indels   54;  Gaps   11;

Qy        971 AGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATA 1030
              | | | |||||||| ||||   ||||| |||||||||  || || |  |    |||||
Db     100972 ACAAAATACAGGCACACCTCAAAGATAATGTGGGTTTCATTTCAGATAATGGAAATAAAG 100913

Qy       1031 CAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCA 1090
                 |||||            ||  | |  ||| || ||||| || |   |||| ||||||
Db     100912 TGAATAT----------TGTAATAAAGCAAATTGTGTCACGTAAATTGCTTGGTTTCCCA 100863

Qy       1091 GTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTAT 1150
              |||||||||||||||  |   || |||||||||||||| |||| |||||||||  ||||
Db     100862 GTGCATATAAAAGTTATATCTACACTACACTGTAGTCTATTAAATGTGCAATAGAATTAT 100803

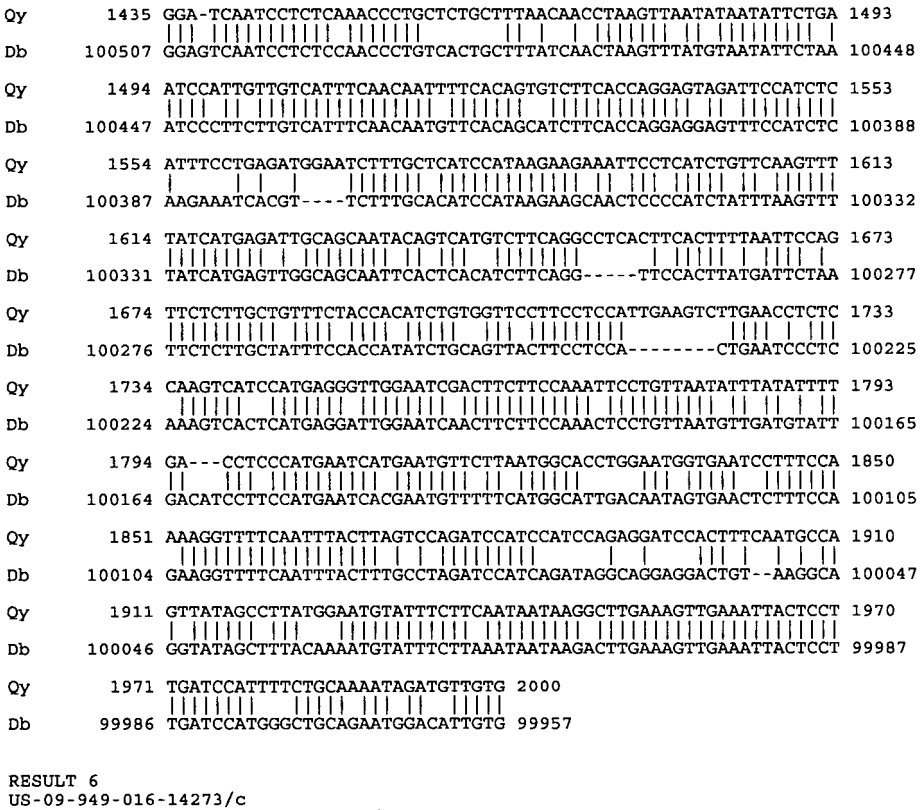
Qy       1151 GTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC 1210
              ||||||| |          | ||||||  |  || |||||||||||||| ||||||||||
Db     100802 ATCTAAAATA----TGTACATAATTTAAGCATACTTTATTACTAAAAAA-GCTAACAATC 100748

Qy       1211 ATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGA 1270
              || ||||| || |||||||||||||||| ||||||||||  || |||   | | || |||
Db     100747 ATCTGAGCCTTTAGTGAGTTGTAATCTTGTTGCTGGTGGAGGGGCTTGCTTCAATGTTGA 100688

Qy       1271 CTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTT 1319
                | |            ||| |   | |||||||| || | |||||||||||| |||
Db     100687 TGGCTGCTGACTGATCAGGGTGGTGGCTGCTGAAGGTTGGTGTGGCTGTGGCAATTTTAA 100628

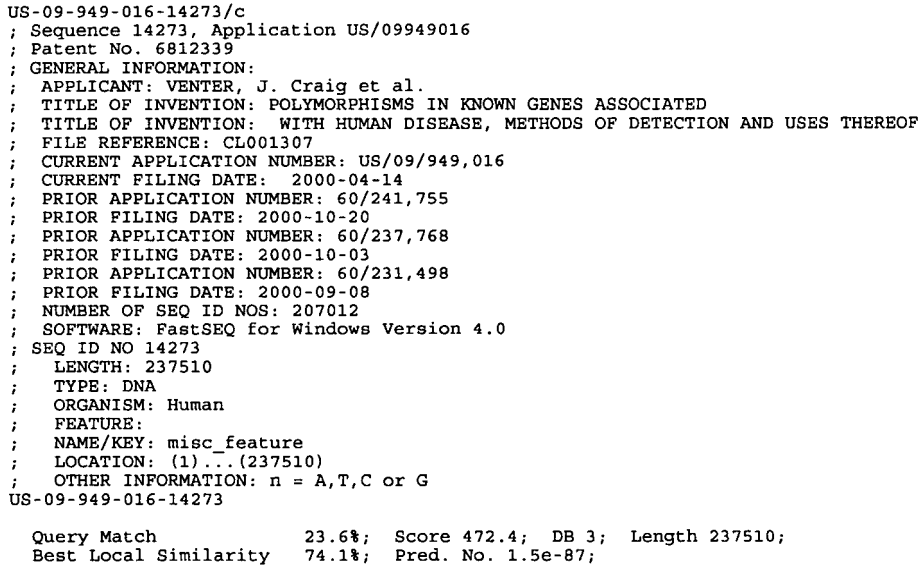
Qy       1320 AAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTC 1374
              |||     ||||||  |||| |||| | |||| |||||||  |   ||| |||||| |||||
Db     100627 AAAAGAAGACAACAACGAAGTTTGCCACATCAATTGACTCCTTTTTTTTATGAAAAATTTC 100568

Qy       1375 TCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATT 1434
              |||  ||  || |||||| |||||||||||||||  ||  |||||||||||||| ||||||
Db     100567 TCTGAAGGCTGCGATGCTGTTTGATAGCATTTTACCCATGGTAGAACTTCTTTCAAAATT 100508
```

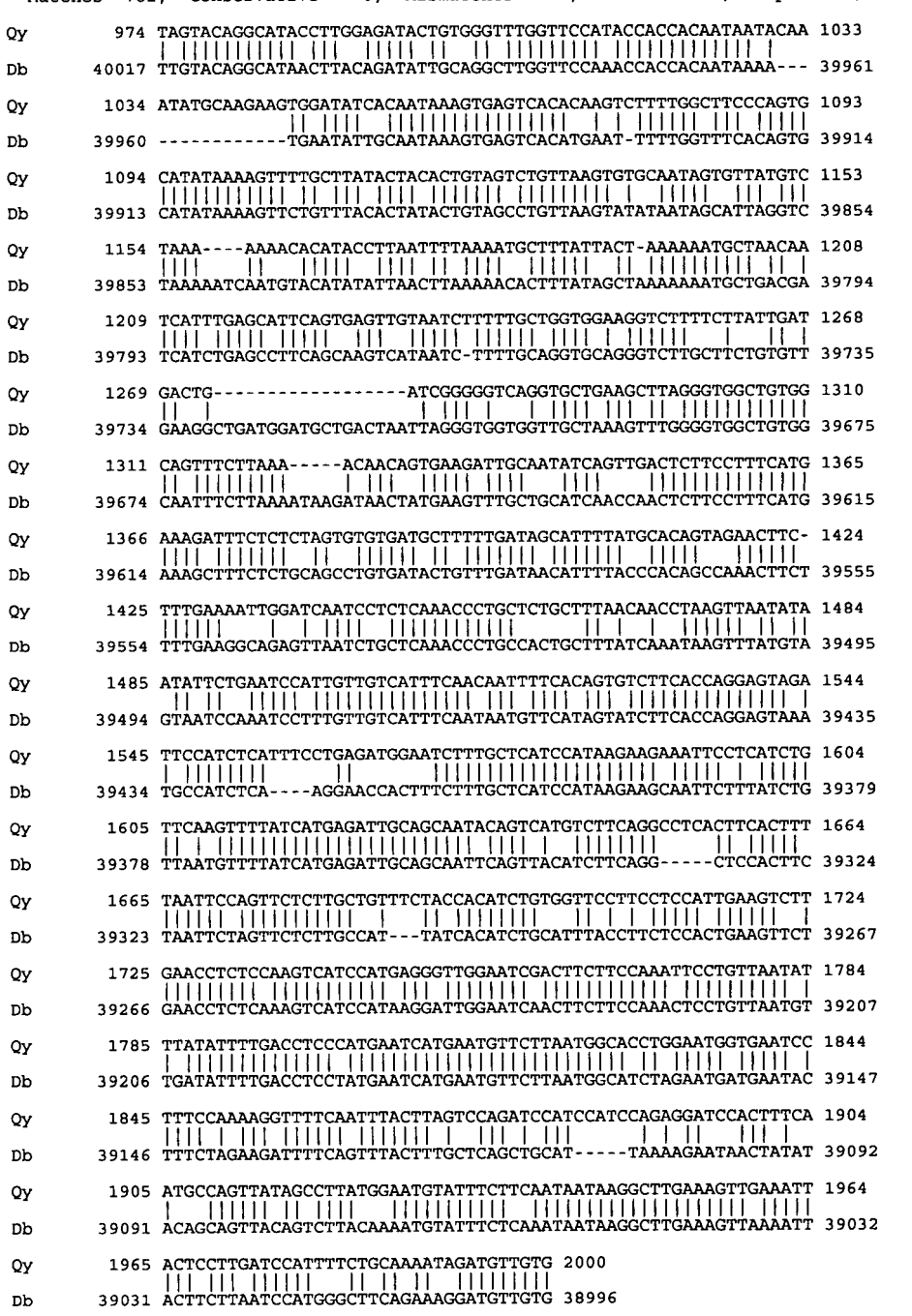


RESULT 6
US-09-949-016-14273/c
; Sequence 14273, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14273
; LENGTH: 237510
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(237510)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14273

Query Match 23.6%; Score 472.4; DB 3; Length 237510;
Best Local Similarity 74.1%; Pred. No. 1.5e-87;



Matches 782; Conservative 0; Mismatches 211; Indels 63; Gaps 12;



```
RESULT 7
US-09-949-016-13173/c
; Sequence 13173, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13173
;   LENGTH: 360470
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13173

  Query Match             23.4%;  Score 468.2;  DB 3;  Length 360470;
  Best Local Similarity   72.6%;  Pred. No. 1.2e-86;
  Matches  762;  Conservative    0;  Mismatches  233;  Indels   54;  Gaps   10;

Qy        975 AGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAA 1034
              | |||||| |||| || |||||| ||  || || |||| | || ||| |||||
Db     253971 ATTACAGGGATACTTTAGAGATATTGCAGGGTTAGTTCAAGACTACCGCAATA------- 253919

Qy       1035 TATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGC 1094
                || ||  |||  || || ||||| | |||| | | || ||| |  |||||||
Db     253918 --------AACTGAGTATTGCAGTACAGTGATTTACACCAATTTTCTGGTTAGCCAGTGC 253867

Qy       1095 ATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCT 1154
              || ||| |||| || |||  |||  ||||||||| ||||||||| |||||| ||||||
Db     253866 ATGTAAGAGTTATGTTTACGCTATTCTGTAGTCTATTAAGTGTGAAATAGTATTATGTGA 253807

Qy       1155 AAAAAAACACATA-----CCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAAT 1209
              ||| ||  |  ||     | | | ||  ||||| ||||||| ||||||||||||||||
Db     253806 AAATAACAATGTACATACCTTAATTTAAAAAATACTTTATTGCTAAAAAATGCTAACAAT 253747

Qy       1210 CATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATG 1269
               || ||||  ||||| |||||||||  |||||||||||||| ||||||  ||   || ||
Db     253746 TATCTGAGTCTTCAGCGAGTTGTAACGTTTTTGCTGGTGGAGGGTCTTGCCTCCATGTTG 253687

Qy       1270 ACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCT 1318
              |  | |            ||| |  || ||||   |  | ||||||||||||| |||||
Db     253686 ATGGCTGCTGACTGATCTGGGTGGTAGTTGCTTGTGGCTGGGGTGGCTGTGGCAATTTCT 253627

Qy       1319 TAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTT 1373
              ||||     |||| | ||||| |||| |||    ||||||||| | ||||||||||| |
Db     253626 TAAAATAAGACAATAATGAAGTTTGCTATACTGATTGACTCTTTCCTTCATGAAAGATAT 253567

Qy       1374 CTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAAT 1433
              |||| |||  ||  ||||| || || ||||||| || ||||||||||||||||||||||
Db     253566 CTCTGTAGCATGCAATGCTGTTGGACAGCATTTAATCCACAGTAGAACTTCTTTGAAAAT 253507

Qy       1434 TGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTG 1492
              |||| |||||||||| ||| ||||       || || | ||||||| || |||||||||
Db     253506 TGGAGTCAATCCTCTTAAATCCTGTCACTACTTTACCAACTAAGTTTATGTAATATTCTA 253447

Qy       1493 AATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCT 1552
              ||||| | ||||||||     |||  | |     | |    |  ||||||||||||
Db     253446 AATCCTTGGTTGTCATCATTTCAACAATGTCTACAGCATTTTTATCAGTAGATTCCATC- 253388

Qy       1553 CATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTT 1612
                 ||    ||       ||||| |||||||||| |||| || |||| ||||||||||||

Db     253387 ---TTAAGAAGCCCCTTTCTTTACTCATCCATAGGAAGTAACTCCTTATCTGTTCAAGTT 253331

Qy       1613 TTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCA 1672
              | |||||||||||||||||| |||||| |||||||       ||||  | |||||| |
Db     253330 TCATCATGAGATTGCAGCAATTCAGTCACATCTTCAGG-----TTTCATCTCTAATTCTA 253276

Qy       1673 GTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCT 1732
              || | |||||| |||  || ||||| || ||| |||| |||| |||||   |||||  ||
Db     253275 GTCCCCTTGCTATTTTGACGACATC-GTAGTTACTTCTTCCACTGAAGCAGTGAACTACT 253217

Qy       1733 CCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTT 1792
                |||||||||||| |||||||||| |||||||    || |||||| ||| || ||||||
Db     253216 TAAAGTCATCCATGAAGGTTGGAATCAACTTCTTTTCAACTCCTGTCAATGTTGATATTT 253157

Qy       1793 --TGACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCA 1850
                |||||| |||||||| |||| |||||||| || || ||||||||||||||||||||
Db     253156 GACCTCCTCCCGTGAATCATAAATGCTCTTAATGACATCTAGAATGGTGAATCCTTTCCA 253097

Qy       1851 AAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCA 1910
              || |||||||||||||| |  ||||||||      | | ||||  |||| || ||| ||
Db     253096 GAAAGTTTTCAATTTACTTTGCTCAGATCCA-----CTATAGGAATCACTATCTATGGCA 253042

Qy       1911 GTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCT 1970
                |||||||||||  ||||||| |||| ||||||||  ||||||  ||||||||||||||
Db     253041 AGTATAGCCTTATAAAATGTATGTCTTAAATAATAACACTTGAATATTGAAATTACTCCT 252982

Qy       1971 TGATCCATTTTCTGCAAAATAGATGTTGT 1999
              ||||| ||   || || |   ||||||||
Db     252981 TGATCTATGGACTACAGAGGGGATGTTGT 252953


RESULT 8
US-09-949-016-15390
; Sequence 15390, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15390
;   LENGTH: 235064
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(235064)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15390

  Query Match             23.3%;  Score 465.8;  DB 3;  Length 235064;
  Best Local Similarity   73.1%;  Pred. No. 3.5e-86;
  Matches  779;  Conservative    0;  Mismatches  232;  Indels   54;  Gaps   12;

Qy        972 GATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATAC 1031
              | || |||||| |||||||||||||||||| || ||||||||||  | ||| ||||||  |
Db      47772 GCTAATACAGGAATACCTTGGAGATACTGAGGATTTGGTTCCAGGCTACCCCAATAAAGC 47831

Qy       1032 AAATATGCAAGAAGTGG----------ATATCACAATAAAGTGAGTCACACAAGTCTTTT 1081
              ||||||   |  |  |            ||||  |||||||| ||||||||  |   ||||
Db      47832 AAATATTGCAATAAAGCAGTAAATCAAATATTTCAATAAAGCGAGTCACATGAACTTTTT 47891

Qy       1082 GGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAA 1141
```

```
Db      47892 GGTTTCCCAGTGCGTATAAAAGTTATGTTTACACTATGCTGTAGTCTATTAAGTGTGAAA 47951

Qy       1142 TAGTGTTATGTCTAAAAAAA---CACATACCTTAATTTTAAAATGCTTTATTACTAAAAA 1198

Db      47952 TAGCATTATGTCTAAAAACAATGTACATACCTTAATTAAAACATACTTCATTGCTAAAAA 48011

Qy       1199 ATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTT 1258

Db      48012 ATGCTAACGATCATCTGAACCTTCAGCAAGCCAT-ATCTTTTTGCTGGTGGAGGGTCTTG 48070

Qy       1259 TCTTATTGATGACTGATCGGGGGTCA-----------GGTGCTGAAGCTTAGGGTGGCTG 1307

Db      48071 CCTCCATGTCAGTGGCTGCTGACTGATCAGATTGGTGGTTACTGCAGGTTGGAGTGATTG 48130

Qy       1308 TGGCAGTTTCTTAAAACAACA----GTGAAGATTGCAATATCAGTTGACTCTTCCTTTCA 1363

Db      48131 TGACAATTTCTTAAAATAAGACAACATGAAGTTTGACATATCAGTTGACTCTTC--TTTC 48188

Qy       1364 TGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTT 1423

Db      48189 GTGAATGATTCTCTGTATCATGCGATGCTGTTTGATAGCATTTTACCCAGAGTAGAACTT 48248

Qy       1424 CTTTGAAAATTGGA-----TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTT 1478

Db      48249 CTTTCAAAATTGGAGTCCGTCCTTCCCCTCAAACCCTGCCACTGTTTTATCAACTAAGTT 48308

Qy       1479 AATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGG 1538

Db      48309 --TATGATGGCGTAAATTTTTTGTTGTCATTTCAACAGTGTTTATAGCTTCTTCACCAGG 48366

Qy       1539 AGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCT 1598

Db      48367 AGTAGATTCTATCTCAAAAAACCACTT----TTCTTGATTAACTGTAAAAAGTAACTCCT 48422

Qy       1599 CATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTT 1658

Db      48423 CATCCAGTCAAGTTTGATCATGAGATCACAGCAATTCAGTCACATCTTCAGG-----CTC 48477

Qy       1659 CACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGA 1718

Db      48478 CTCTTATAATTCTAATTCTCTTGCTATTAAAACCATGTCTGCAGTTACTTCCTCCACTGA 48537

Qy       1719 AGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGT 1778

Db      48538 AGTCTCGAACCCCTCAAAGTCATCCATGAGGGTTGGAATCAGCTTCTTCCAAAGTCCTGT 48597

Qy       1779 TAATATTTATATTTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAAT 1835

Db      48598 TAATGTTGATATTTTTATCTACTCCCATGAATCATGAATGTTCTGAATGGCATTCAGAAT 48657

Qy       1836 GGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGAT 1895

Db      48658 GGCGAATCCTTTCCAGAATGTTTTCCATTGATTTTGCCCAG----ATCCACCAGAAGAAT 48713

Qy       1896 CCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAA 1955

Db      48714 CAATTATCTGTGGCAACTCTAGGCTTATGAAATGTATTTCTTAAATAATAAAACTTGAAA 48773

Qy       1956 GTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000

Db      48774 GTTGAAATGCCTCCTTCATCCATGGGCTACAGAATGGATAATGTG 48818


RESULT 9
US-09-949-016-17422
; Sequence 17422, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17422
;   LENGTH: 164061
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17422

  Query Match             23.3%;  Score 465.6;  DB 3;  Length 164061;
  Best Local Similarity   74.9%;  Pred. No. 3.5e-86;
  Matches  789;  Conservative    0;  Mismatches  204;  Indels   61;  Gaps   14;

Qy        975 AGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAA 1034

Db      75766 ATTATAGGCATATCTTGGAGATATTGCAGGTTTGGTTCCAGACCACTGCAGT-------- 75817

Qy       1035 TATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGC 1094

Db      75818 -------GAAGTGAATGACACAATAAAGCAAGTCACAAATTTTATTTCCTTTGCATGTGC 75870

Qy       1095 ATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAG----TGTGCAATAGTGTTAT 1150

Db      75871 ATATAAAAGTTATCTTTCCACTTTACTGTAGCCTCTTAAGGTACAGTGGCATTATTTCTT 75930

Qy       1151 GTCTAAAAAAACACATACCTTAA-TTTTAAAATGCTTTATTAC-TAAAAAATGCTAACAA 1208

Db      75931 TAAAAAAAATGCATATACCTTAACTTAAAAAATGCTTTATTGCTTAAAAAATGCTAATGA 75990

Qy       1209 TCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCT---TATT 1265

Db      75991 TCATCTGAGCCTTCAGTGAGTTATACTCTTTTTGCTAGTGAAGGGTCTTGCCTCAGTATT 76050

Qy       1266 GATGACTGATCGGGGGTCAGG--------TGCTGAAGCTTAGGGTGGCTGTGGCAGTTTC 1317

Db      76051 GATGGCTGCTGGCTGATAAGGATGGTGATTGCTGAAAGTTGGGGTGGCTGTGACAGTTTC 76110

Qy       1318 TTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATT 1372

Db      76111 TTAAAATAAGACAAAAATGAAATTTGCTGCATCAATGAGTTCTTCCTTTCACAAAAGATT 76170

Qy       1373 TCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAA 1432

Db      76171 TATCTGTGGTGTGCAATGCTGTCTGATTGCATTTGACCCACAATAGAACTTCTTTCAAAA 76230

Qy       1433 TTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCT 1491

Db      76231 TTGGAGTCAATCCTCTCAAACCCTGCTGCTGCTTTATCAGCTAAGTTTATGTAATATTCT 76290

Qy       1492 GAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATC 1551

Db      76291 AAATCCTTTGTTGTCATTTCAACAATGTTCACAGCATCTTCATCAGGAGCAGTTTCCATC 76350

Qy       1552 TCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGT 1611

Db      76351 TCAAGAAACCACTT----TATTTGCTCATTCATAAGAAGCAATTTCTCATCCATTCAAGT 76406

Qy       1612 TTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCC 1671

Db      76407 TTTATCATGAGAGTGTAGCAGTTCAGTCACATCTTCAGG-----CTCTACTTCTAATTCT 76461

Qy       1672 AGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTC--TTGAACC 1729

Db      76462 AGTTCTCTTGCTATTTCTACCACATCTGCAGTTAATTTCTTCTTCTGAAGCCTTTAACAC 76521

Qy       1730 TCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATA 1789
```

```
Db      76522 TTTCCACATCATCCATGAGCATTGGAATCAACTTCTTCCAAACTCTTGTTTATGTTGATA 76581

Qy       1790 TTTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTT 1846

Db      76582 TTTTGATCTCTTCCCATGAATCATGAATGTTCTTAATGGCATCTAGAATGGTGAATCCTT 76641

Qy       1847 TCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAAT 1906

Db      76642 TCCAGAAGGTTTTCAATTTACTTTGCCCAGATCCAT-----CAGAGGAATCACTGTCTGT 76696

Qy       1907 GCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAG----GCTTGAAAGTTGAAA 1962

Db      76697 GGCAACTATAGCCTTACAAGATGTGTTTCTTAAATAATAAGATTTTCTTGAAAGTCTAAA 76756

Qy       1963 TTACTCCTTGATCCATTTTCTGCAAAATAGATGT 1996

Db      76757 TGACTCCTTGATTCATAGGCTACAAAATGGATAT 76790


RESULT 10
US-09-949-016-11869/c
; Sequence 11869, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11869
;   LENGTH: 194714
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-11869

  Query Match            23.1%;  Score 462.6;  DB 3;  Length 194714;
  Best Local Similarity   74.3%;  Pred. No. 1.5e-85;
  Matches  778;  Conservative    0;  Mismatches  209;  Indels   60;  Gaps   13;

Qy        979 CAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATG 1038

Db      62690 CAGGTGTACCTAGGAGATATTACAGGTTTGGTTCTAGACCACCACAATAAAGCAAGTAT- 62632

Qy       1039 CAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATAT 1098

Db      62631 -----------TATGATAAAAAAGT------CACACAAATTTTTGTTTCCAAGTGCATAT 62589

Qy       1099 AAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAA- 1157

Db      62588 AAAAGTTATATTTATACTATACTGTAGTCTGTTAAGTGTACAATAGCATTATGTCTAAAA 62529

Qy       1158 ----AAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213

Db      62528 TCACAATTCATATACCTTAATTTAAAAATACTTCGTTGCTAGAAAATGCTAAGGATCATC 62469

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTG 1273

Db      62468 TGAGCCTCCAGTGAGTTGTGATCTTTTTGCTGGAGGAGGATCTCGCCTCCACGTTGATGG 62409

Qy       1274 AT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCT---- 1318

Db      62408 CTGCTGACTGATCAGGGTAGTAGTTGCTGAAGAATGGGGTGACTATGGCGATTTCTTAAA 62349

Qy       1319 -TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCT 1377

Db      62348 ATAAAACAACAATGAAGTTTGCCACAACAATTGACTCTTCCTTTCATGAAAGATTTCTCT 62289

Qy       1378 CTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA 1437

Db      62288 GTAGCATGCTACATTATTTGATAGCATTTTACCCACAACAGAACTTCCTTCAAAATTGGA 62229

Qy       1438 -TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATC 1496

Db      62228 GGCAATCCTCTCAAACCCTACTACTGCTTTATCAACTAAGTTTATGTAATATTCTAAATT 62169

Qy       1497 CATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATT 1556

Db      62168 CTTTTTCGTTATTTCGACAATGTT-ATGGCATCTTCATTAGGAATAGATTTCATCTCAAG 62110

Qy       1557 TCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTAT 1616

Db      62109 AAAACACTT----TCTTTGCTCATCTGTAAGAATCAACTCCTCATTTCCTCAAGTTTTAT 62054

Qy       1617 CATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTC 1676

Db      62053 CATGAGTTTG-AGCAATGCAGTCACATCTTCAGG-----CTTTATTTCTAATTCCAGATC 62000

Qy       1677 TCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAA 1736

Db      61999 TCATACTGTTTCCACCACATCTGTAGTTAACTCCTCCACTGAAGTCTTGAATTCCTTAAA 61940

Qy       1737 GTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC 1796

Db      61939 GTCATCCATGAGGGCCGGAATAAACTTTTTCTAAACTTTTGTTAATGTTAATATTTTGAC 61880

Qy       1797 CTC---CCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAA 1853

Db      61879 CTCCTTCCATGAATCACGAATGTTCTTAATGACATCTAAAAGGGCAAATTCTTTCCAGAA 61820

Qy       1854 GGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTT 1913

Db      61819 GG-TTTCAATTTACTTTGCCTACATCCAT-----CAGAGGAATCATTACATATGGCAGCT 61766

Qy       1914 ATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGA 1973

Db      61765 ATAGCCTTATGAAATGTATTTCTTAAATAATAAGACTTGAAAGTTACAATTTTTCCTTGA 61706

Qy       1974 TCCATTTTCTGCAAAATAGATGTTGTG 2000

Db      61705 TCCATAGGCTGTTGAATGAATATTGTG 61679


RESULT 11
US-09-949-016-15474/c
; Sequence 15474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15474
;   LENGTH: 196714
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15474
```

```
  Query Match             23.1%;  Score 462.6;  DB 3;  Length 196714;
  Best Local Similarity   74.3%;  Pred. No. 1.5e-85;
  Matches  778;  Conservative    0;  Mismatches  209;  Indels   60;  Gaps   13;

Qy        979 CAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATG 1038
              ||||  ||||| ||||||| |   |||||||||| | ||||||||||||  ||| |||
Db     112690 CAGGTGTACCTAGGAGATATTACAGGTTTGGTTCTAGACCACCACAATAAAGCAAGTAT- 112632

Qy       1039 CAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATAT 1098
                      ||| | || |||||      ||||    |||| | |||| |||||||||
Db     112631 -----------TATGATAAAAAAGT------CACACAAATTTTTGTTTCCAAGTGCATAT 112589

Qy       1099 AAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAA- 1157
              ||||||| |  |||||||| |||||||||||||||||| ||||||   ||||||||||
Db     112588 AAAAGTTATATTTATACTATACTGTAGTCTGTTAAGTGTACAATAGCATTATGTCTAAAA 112529

Qy       1158 ----AAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATT 1213
                  ||  || |||||||||||| ||||| ||| || ||| ||||||||||  |||||
Db     112528 TCACAATTCATATACCTTAATTTAAAAATACTTCGTTGCTAGAAAATGCTAAGGATCATC 112469

Qy       1214 TGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTG 1273
              ||||| | |||||||||| |||||||||||||| ||| | |||    ||     | |||  |
Db     112468 TGAGCCTCCAGTGAGTTGTGATCTTTTTGCTGGAGGAGGATCTCGCCTCCACGTTGATGG 112409

Qy       1274 AT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCT---- 1318
                |            |||     || ||||||||  | ||||| || ||||  |||||
Db     112408 CTGCTGACTGATCAGGGTAGTAGTTGCTGAAGAATGGGGTGACTATGGCGATTTCTTAAA 112349

Qy       1319 -TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCT 1377
               |||||||||| ||||| |||| | | || ||||||||||||||||||||||||||||||
Db     112348 ATAAAACAACAATGAAGTTTGCCACAACAATTGACTCTTCCTTTCATGAAAGATTTCTCT 112289

Qy       1378 CTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA 1437
              |||  ||  |   | ||||||||||||||  ||||  |||||||| || |||||||||
Db     112288 GTAGCATGCTACATTATTTGATAGCATTTTACCCACAACAGAACTTCCTTCAAAATTGGA 112229

Qy       1438 -TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATC 1496
               |||||||||||||||| ||      || |  | |||||| || ||||||||| |||
Db     112228 GGCAATCCTCTCAAACCCTACTACTGCTTTATCAACTAAGTTTATGTAATATTCTAAATT 112169

Qy       1497 CATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATT 1556
              | || | || ||||| ||||| || |  |  ||||||  |||| |||||| |||||||
Db     112168 CTTTTTCGTTATTTCGACAATGTT-ATGGCATCTTCATTAGGAATAGATTTCATCTCAAG 112110

Qy       1557 TCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTAT 1616
                |  |   |||||||||||||  ||||||  || |||||||  |  ||||||||||
Db     112109 AAAACACTT----TCTTTGCTCATCTGTAAGAATCAACTCCTCATTTCCTCAAGTTTTAT 112054

Qy       1617 CATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTC 1676
              ||||||  ||| |||||| ||||||  |||||||||      ||| | || |||||||| ||
Db     112053 CATGAGTTTG-AGCAATGCAGTCACATCTTCAGG-----CTTTATTTCTAATTCCAGATC 112000

Qy       1677 TCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAA 1736
              || | |||||||  |||||||||||  |||   |||||||  ||||||||||||   ||  ||
Db     111999 TCATACTGTTTCCACCACATCTGTAGTTAACTCCTCCACTGAAGTCTTGAATTCCTTAAA 111940

Qy       1737 GTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC 1796
              ||||||||||||||  |||||  |||| ||| ||| |  |||||||  || |||||||||
Db     111939 GTCATCCATGAGGGCCGGAATAAACTTTTTCTAAACTTTTGTTAATGTTAATATTTTGAC 111880

Qy       1797 CTC---CCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAA 1853
              |||    |||||||||| |||||||||||||| || ||  || ||  ||| |||||||  ||
Db     111879 CTCCTTCCATGAATCACGAATGTTCTTAATGACATCTAAAAGGGCAAATTCTTTCCAGAA 111820

Qy       1854 GGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTT 1913
              || ||||||||||||| | | | ||||||     |||||||  || |    ||| ||| |
Db     111819 GG-TTTCAATTTACTTTGCCTACATCCAT-----CAGAGGAATCATTACATATGGCAGCT 111766

Qy       1914 ATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGA 1973
              |||||||||||  |||||||||||  |||||||||  ||||||||||   ||||  |||||||
Db     111765 ATAGCCTTATGAAATGTATTTCTTAAATAATAAGACTTGAAAGTTACAATTTTTCCTTGA 111706

Qy       1974 TCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||||   |||   |||  || |||||
Db     111705 TCCATAGGCTGTTGAATGAATATTGTG 111679


RESULT 12
US-09-949-016-14179
; Sequence 14179, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14179
;   LENGTH: 340380
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(340380)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14179

  Query Match             23.1%;  Score 461.2;  DB 3;  Length 340380;
  Best Local Similarity   73.2%;  Pred. No. 3.3e-85;
  Matches  779;  Conservative    0;  Mismatches  218;  Indels   67;  Gaps   12;

Qy        956 TTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCAT 1015
              ||| | ||  |   |   |||||||||||||| |||||||  || ||  || |  ||||||
Db      34227 TTTATTTATTGTCTTTTCTAGTACAGGCATATCTTGGAGTTATTGCAGGCTCAGTTCCAG 34286

Qy       1016 ACCACCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAG 1075
              | |||  |||||                |||||  ||||||||||||||  |||||||||
Db      34287 ATCACTGCAATA--------------AAGTGAATATCACAATAAAGCAAGTCACACACA 34331

Qy       1076 TCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGT 1135
              | |||||  |||| ||||||||  |||||||  || |||||  || ||||||||||||||||
Db      34332 TTTTTTGT-TTCCTAGTGCATAGAAAAGTTATGTTTATATTATACTGTAGTCTGTTAAGT 34390

Qy       1136 GTGCAATAGTGTTATGTCTAAAAAAA--CACATACCTTAATTTTAAAATGCTTTATTACT 1193
              | |||||||  ||||||||  ||||||    |  ||||||||||  |||||  | |||||  ||
Db      34391 GAGCAATAGCATTATGTCTTAAAAAATGTATCTACCTTAATTAAAAAATACATTATTGCT 34450

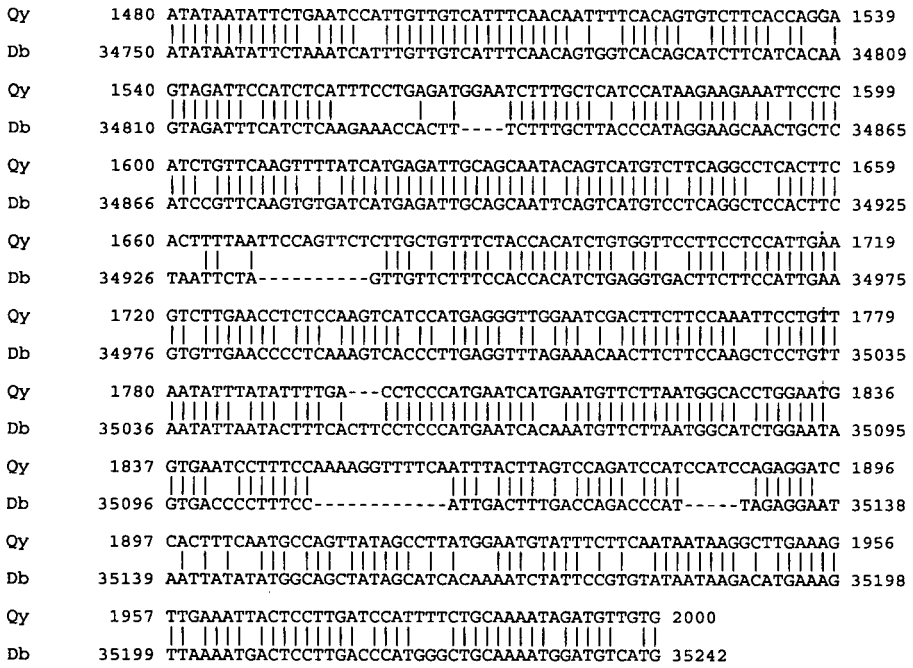
Qy       1194 AAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGG 1253
               ||||  | ||  |  |||||  ||||  ||||    ||||||||||||  ||||| ||
Db      34451 -AAAAGTACTGATGATCATCTGAGCCTTCAATGAGGCATAATCTTTTTGCTAGTGGAGGG 34509

Qy       1254 TCTTTTCTTATTGATGACTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGT 1302
              |||||  || | |   ||  | |            |||  |   ||  | ||||||  || ||||
Db      34510 TCTTTCCTCAATATGGATGGTTGCTGACTAATCAGGGTGGTAGTTTCTGAAGGTTTGGGT 34569

Qy       1303 GGCTGTGGCAGTTTCTTAAAACA--ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT 1360
              | ||||||||  |||| |||||  |  |||  ||||  ||||  | ||||||||||  | | ||
Db      34570 GACTGTGGCAATTTCCTAAAATAAGACAATGAAATTTGCCACATCAGTTGACTTTGCTTT 34629

Qy       1361 TCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAA 1420
              ||   ||||||||||||  |||   || ||||||  || ||||||||||||   ||  |||||||
Db      34630 TCGCAAAAGATTTCTCTATAGCATGGGATGCTGTTGGATAGCATTTTACCCATAGTAGAA 34689

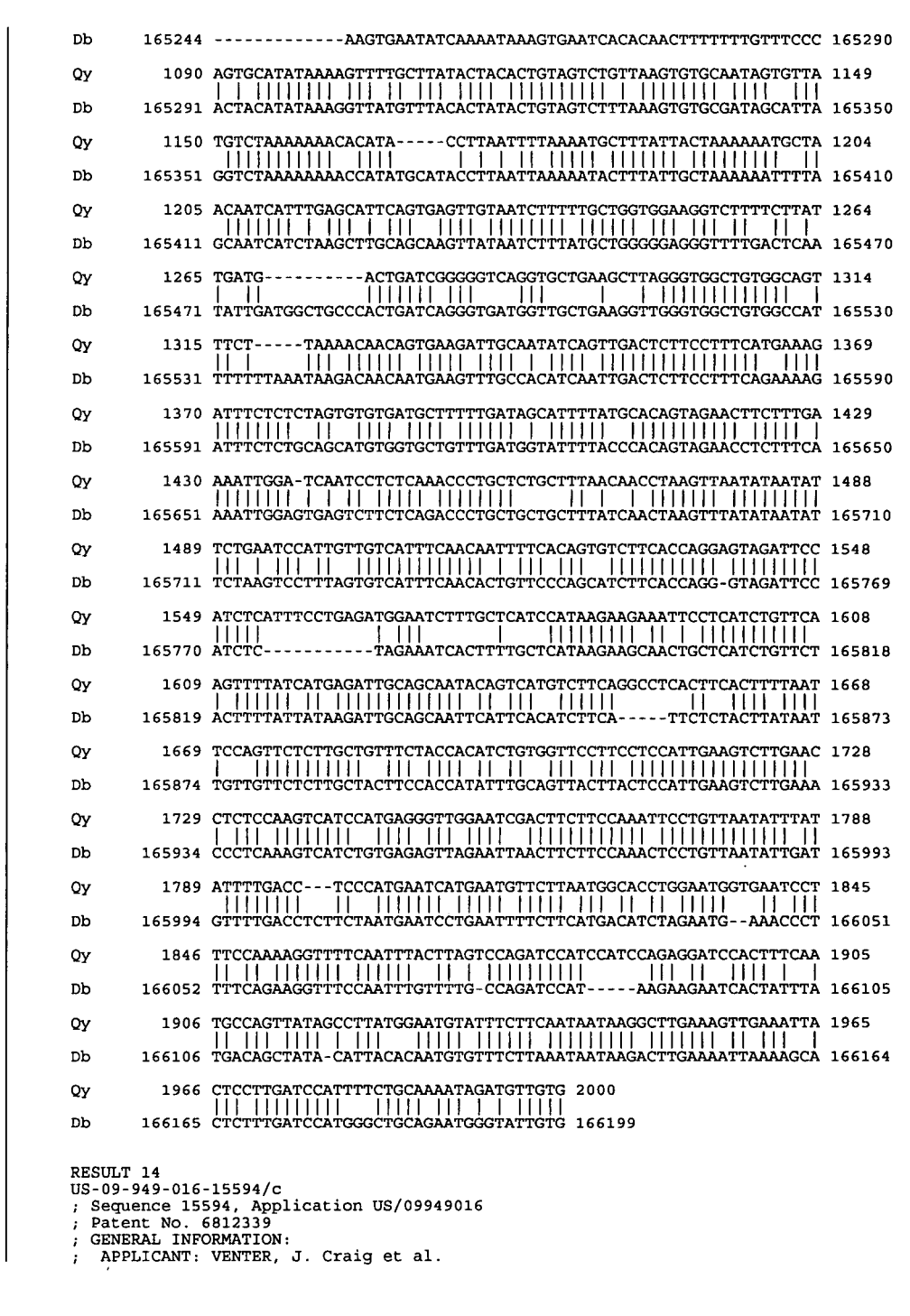
Qy       1421 CTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTA 1479
              |||||||  |||||||||  ||||||||||||||||  ||         | |  | |||  ||||
Db      34690 CTTCTTTCAAAATTGGAGTCAATCCTCTCAAACCTTGTCAGTGCGTTATCAACTATGTTA 34749
```



RESULT 13
US-09-949-016-13358
; Sequence 13358, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13358
; LENGTH: 260247
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13358

Query Match 22.7%; Score 454.2; DB 3; Length 260247;
Best Local Similarity 73.6%; Pred. No. 8.6e-84;
Matches 777; Conservative 0; Mismatches 213; Indels 65; Gaps 13;





RESULT 14
US-09-949-016-15594/c
; Sequence 15594, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

```
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15594
;   LENGTH: 59948
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15594

  Query Match             22.7%;  Score 453.8;  DB 3;  Length 59948;
  Best Local Similarity   72.9%;  Pred. No. 7.6e-84;
  Matches  772;  Conservative    0;  Mismatches  222;  Indels   65;  Gaps   12;

Qy        971 AGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATA 1030
              | |  |||||||||| |||  |||||| |||||||| | ||||| |  |||||||| |
Db      53894 AAAATGTACAGGCATGCCTGAGAGATATTGTGGGTTAGCTTCCAGAATACCACAATGAAG 53835

Qy       1031 CAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCA 1090
              |||                  |||||||||| ||||||||| |  | |||| | |||| |
Db      53834 CAAG--------------TATCACAATAAAATGAGTCACAAA--TTTTTTTGTTTCCTA 53792

Qy       1091 GTGCATATAAAAGTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTAT 1150
              |||||||||||||| || ||||| ||| ||||||||| ||| || | || |||  ||||
Db      53791 GTGCATATAAAAGTTATGTTTATATTACTCTGTAGTCTATTAGGTATACATTAGCATTAT 53732

Qy       1151 GTCTAAAAAAAC-----ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAA 1205
              |||||| |||||     |||||||||||||  ||||  ||||||| || ||||||||||
Db      53731 GTCTAAGAAAACAATATACATACCTTAATTAAAAAACACTTTATTCCT-AAAAATGCTAA 53673

Qy       1206 CAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATT 1265
              |||||  |||  |||||  |||||  |||||||||||| |||| || |||   ||  |
Db      53672 TGATCATCTCAGCCTTCAGCAAGTTGCCATCTTTTTGCTGATGGAGGGGCTTGCTTTGAT 53613

Qy       1266 GATGACTGATC-----------GGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGT 1314
              | |||     ||           ||  | | | |||||||| || ||||||||||||| | |
Db      53612 GTTGATATCTCCTGACTGATCAGGACGGCGGTTGCTGAAGATTGGGGTGGCTGTGGAAAT 53553

Qy       1315 TTCTTAAAACA--ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATT 1372
              |||  |||||  |   |||  | | ||          || |||||||| |||| ||||||||||
Db      53552 TTCCTAAAATAAGACAATAATGAAATTTGCCACATTTGACTCTTTCTTTTATGAAAGATT 53493

Qy       1373 TCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAA 1432
              |||||  |||  |||||||||  |||||  ||||  ||||  |||||  ||||||||||  ||||
Db      53492 TCTCTGTAGCATGTGATGCTGTTTGACAGCACTTTACCCACAGGAGAACTTCTTCCAAAA 53433

Qy       1433 TTG-GATCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCT 1491
               ||   |||||||||||||  |||||      || |  | |||  ||  || |||| ||||
Db      53432 ATGTTGTCAATCCTCTCAAATTCTGCTGCTGCTTTATCAACTACATTTATGTAATCTTCT 53373

Qy       1492 GAATCC-------ATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGA 1544
               |||||        |||| ||||||| |||| | ||||      |||| |  |||||||||
Db      53372 AAATCCTTTATGGTTTGTGGTCATTTTAACACTGTTCATGACTTCTTTATTAGGAGTAGA 53313

Qy       1545 TTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTG 1604
              ||||||||||      | || |     | ||||||||||  |||||||  || |||| |  ||
Db      53312 TTCCATCTCAAGAAATCAGTT----TATTTGCTCATCTTTAAGAAGCAACTCCTTACTTG 53257

Qy       1605 TTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTT 1664
              |||||||||||||||||||||||||||| | ||||||  ||||||||      || | |||
Db      53256 TTCAAGTTTTATCATGAGATTGCAGCAGTTCAGTCACATCTTCAGG-----CTCCCCTTC 53202

Qy       1665 TAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTT 1724
              ||||||  |||| ||||||| |||  |  | |||||  ||| |||||||| |||||||||
Db      53201 TAATTCTAGTTTTCTTGCTATTTTCATGATATCTGCAGTTACTTCCTCCAATGAAGTCTT 53142

Qy       1725 GAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATAT 1784
              ||||  ||| ||||||||||||||| ||||||  | ||||  |||||||  || |||||||  |
Db      53141 AAACCCCTCAAAGTCATCCATGAGGATTGGAACCAACTTATTCCAAACTCATGTTAATGT 53082

Qy       1785 TTATATTTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAA 1841
              | |||||||||   | ||||||  |||||||||||||||||||  |||| ||  |||||||||
Db      53081 TGATATTTTGAGCTCTTCCCATAAATCATGAATGTTCTTAAGGGCATCTAAAATGGTGAA 53022

Qy       1842 TCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTT 1901
              |||||||||  | |||| |||||||||      ||||||||||          |  || | |
Db      53021 TCCTTTCCAGGAAGTTTCCAATTTACTATTCTCAGATCCATC---------AAAGCAATC 52971

Qy       1902 TCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAA 1961
              |  || ||| ||||| || |   | |||||||||| ||||| |||   |||||||  ||
Db      52970 ACTGTGGCAGCTATAGTCTCACAAACTGTATTTCTTAAATAACAAGAAATGAAAGTCTAA 52911

Qy       1962 ATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              || | ||||||||||||   ||||| ||| |||| ||||
Db      52910 ATGATTCCTTGATCCATGGGCTGCAGAATGGATGCTGTG 52872


RESULT 15
US-09-949-002-805/c
; Sequence 805, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 805
;   LENGTH: 135687
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(135687)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-002-805

  Query Match             22.7%;  Score 453.4;  DB 3;  Length 135687;
  Best Local Similarity   74.7%;  Pred. No. 1.1e-83;
  Matches  728;  Conservative    0;  Mismatches  206;  Indels   41;  Gaps   11;

Qy       1051 ATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCT 1110
              ||  |||||||| |||| |||||  | |||||| |||||| | ||||||||||| |  |
Db      49232 ATGGCAATAAAGCGAGTTACACAGATATTTTGGTTTCCCAGAGGATATAAAAGTTATTTT 49173

Qy       1111 TATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA----ACACAT 1166
              || |||| |||||| | | ||||||||||||||   ||||||||||||      ||||
Db      49172 TAGACTATACTGTAATTTCTTAAGTGTGCAATAGCACTATGTCTAAAAAACAATGTACAT 49113

Qy       1167 ACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTG 1226
              ||| |||||  ||||| |||||||| || |||||||| |||||||||  | ||  || ||
Db      49112 ACCATAATTAAAAAATACTTTATTGCTGAAAAATGGTAACAATCACCTTAGTCTTTAGCT 49053

Qy       1227 AGTTGTAATCTTTTTGCTGGTGGAAGGTCT----TTTCTTATTGATGACTGATCGGGGGT 1282
              || | ||| ||||||||||||| | |||||     |||  | |||||| |||       | |
Db      49052 AGCTATAAACTTTTTGCTGGTGAAGGGTCTTAACTTTGATGTTGATGGCTGCCGAATGAT 48993

Qy       1283 CAGG--------TGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCT-----TAAAACAACAG 1329
              ||||        |||||| |    |||  | ||||| | |||||      ||| ||||||
Db      48992 CAGGGTGGTGGTTGCTGACGGACGGGGCTGTTGTGGAAATTTCTTAGCATAAGACAACAA 48933
```

```
Qy       1330 TGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGAT 1389
             ||||| |  ||  |    |||||||||||||||||||| ||||| |||| |||  ||||||
Db      48932 TGAAGTTGCCACCACGGATTGACTCTTCCTTTCATGATAGATTCCTCTGTAGCATGTGAT 48873

Qy       1390 GCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTC 1448
             ||| ||||||||||||||  |||||||||||||||||| ||||||||| |||||| | ||
Db      48872 GCTGTTTGATAGCATTTTACCCACAGTAGAACTTCTTTCAAAATTGGAGTCAATCTTTTC 48813

Qy       1449 AAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCAT 1508
             ||  | |||      || |  | ||||||| || ||||||||| |||||  ||| |||| |
Db      48812 AAGGCGTGCCACTGCTTTATCAACTAAGTTTATGTAATATTCTAAATCCTTTGCTGTC-T 48754

Qy       1509 TTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGA 1568
             |||||||||  |||||||  | | || ||||  ||||  |||||||||        | |
Db      48753 TTCAACAATATTCACAGCATTTACAGCAGGGATAGAGTCCATCTCAAGAAACCACTT--- 48697

Qy       1569 ATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCA 1628
              |||||||||||| ||||||||  ||    |||||||| ||||||||||||||||||||
Db      48696 -TCTTTGCTCATCTATAAGAAGCAACAGTTCATCTGTCCAAGTTTTATCATGAGATTGCG 48638

Qy       1629 GCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTC 1688
             |||||  |||| |  |||||  ||     ||  | || |||||| | |||||||| | ||||
Db      48637 GCAATTCAGTTACATCTTCTGG-----CTATATTTCTAATTCTAATTCTCTTGGTATTTC 48583

Qy       1689 TACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAG 1748
              || | |||||  ||| |||||  || ||| ||||  ||||  || ||||||||||||
Db      48582 CACAAGATCTGCAGTTACTTCCAACAGTGATGTCTCAAACCCTTCAAAGTCATCCATGAA 48523

Qy       1749 GGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGA 1805
             || ||||||| |||||||||||||||||||| || || ||| |||||    |||||||| |
Db      48522 GGCTGGAATCAACTTCTTCCAAATTCCTGTTCATGTTGATAATTTGACCTCCTCCCATAA 48463

Qy       1806 ATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTT 1865
             ||||  ||||||| |||||| |  | ||||| ||| |||||||||  | ||||||||||||
Db      48462 ATCACAAATGTTCCTAATGGGATGTAGAATGTTGAGTCCTTTCCAGAGGGTTTTCAATTT 48403

Qy       1866 ACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGG 1925
             |||    ||||| |||     ||||||||  | || |  ||| ||| ||| |||||| |
Db      48402 ACTCTTCCCAGAGCCA-----CCAGAGGAATCTCTATTTATGTCAGCTATGGCCTTACGA 48348

Qy       1926 AATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCC-ATTTTCTG 1984
             |||| |||||||  || ||||||  ||||||||  |||||  | ||||||||||  ||  |||
Db      48347 AATGAATTTCTTAAACAATAAGACTTGAAAGCAGAAATCAGTCCTTGATCCAATGGGCTG 48288

Qy       1985 CAAAATAGATGTTGT 1999
             ||  ||| ||||||||
Db      48287 CAGAATGGATGTTGT 48273
```

```
RESULT 16
US-09-949-016-15698
; Sequence 15698, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15698
;   LENGTH: 145812
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(145812)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15698

  Query Match             22.7%;  Score 453;  DB 3;  Length 145812;
  Best Local Similarity   73.1%;  Pred. No. 1.3e-83;
  Matches  764;  Conservative    0;  Mismatches  225;  Indels   56;  Gaps   12;
```

```
Qy        973 ATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACA 1032
              | | ||| | || |||| ||||||| ||||||||||||||||| || | | || | | ||
Db      43948 AAAATACTGACACACCTCGGAGATATTGTGGGTTTGGTTCCAGACTATCGCAGTGAAGCA 44007

Qy       1033 AATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGT 1092
              |||||                |||||||  | |||||  |  ||| || ||||||||
Db      44008 AATATTG--------------CAATAAAGCAAATCACATGAAGTTTTAGGTTTCCCAGT 44052

Qy       1093 GCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGT 1152
              |||||||||||||| || ||| | ||  |||||||||||||||| ||||||| | ||||||
Db      44053 GCATATAAAAGTTATGTTTAAATTATTCTGTAGTCTGTTAAGTATGCAATAATATTATGT 44112

Qy       1153 CTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCAT 1212
              | |||||  |||||||||||||| ||||| |||||||   |||||   || | ||||||
Db      44113 CCAAAAAAGTACATACCTTAATTTAAAAATACTTTATTGTCAAAAAGCACTCATAATCAT 44172

Qy       1213 TTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTG------ 1266
              |||||  ||||||||||  |||||||||| | |||||  |||||  |||  ||
Db      44173 CTGAGCCTTCAGTGAGTCATAATCTTTTTGTTAGTGGAGAGTCTTGCCTTGATGTTGATG 44232

Qy       1267 ----ATGACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA 1322
                  |||||||  | |||     ||       |   | ||||||||  |||||  | |||||||
Db      44233 GCTACTGACTGACCAGGGTGGTGGCTGCTGAAGGTTGGGTGGCTATGGCAATATCTTAAA 44292

Qy       1323 ACA--ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTA 1380
              | |  ||| ||||| ||||    |||||| ||||||||||||| ||||||| |||||  ||
Db      44293 ATAAGACAATGAAGGTTGCTGCATCAGTGGACTCTTCCTTTCGTGAAAGATGTCTCTGTA 44352

Qy       1381 GTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA--T 1438
              || ||  ||||| |||||||||||||||  | ||||||||||||||||||||||| |||  |
Db      44353 GTATGCAATGCTATTTGATAGCATTTTACCCGCAGTAGAACTTCTTTGAAAATAGGAAGT 44412

Qy       1439 CAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCA 1498
                || | ||||||||||||      ||        ||||| || || |||| | |||||
Db      44413 TGATTCACTCAAACCCTGCCACTGCTT---TGATTAAGTTGATGTAGTATTATAAATCCT 44469

Qy       1499 TTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTC 1558
              |||||||||||||||||||| |||||||  |||||||||||| ||||||||||||||||
Db      44470 TTGTTGTCATTTCAACAATATTCACAGCATCTTCACCAGGAATAGATTCCATCTCAAGAA 44529

Qy       1559 CTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCA 1618
                | |   |||||||||||||||||| ||| ||||||| || ||| ||||||||||
Db      44530 ACCACTT----TCTTTGCTCATCCATAAAAAGCAATTCCTTATATGTCCAAGTTTTATCG 44585

Qy       1619 TGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTC 1678
              ||||||| ||||||| |||||| ||||||| |      || ||||| |||||  |||||||
Db      44586 TGAGATTACAGCAATTCAGTCACGTCTTCAAG-----CTCCACTTCTAATTTTAGTTCTC 44640

Qy       1679 TTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGT 1738
              ||| |  |||||||||||||  ||  |||||| || |||||  || || | ||| |||
Db      44641 TTGTTACTTCTACCACATCTGCAGTGACTTCCTTCACTGAAG-ATTAAATCCCTCAAAGA 44699

Qy       1739 CATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCT 1798
              | ||||||| |   |||||  |||||||||| |||||||||| || | ||||||||||
Db      44700 TACCCATGAGAGCCAGAATCAGCTTCTTCCAAACTCCTGTTAATGTTGACATTTTGACCT 44759

Qy       1799 CC---CATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGG 1855
              ||   ||||||  |  |||||| || ||| |  || ||||||||||  ||||||| |||
Db      44760 CCTCTCATGAACTACAAATGTTTTTCATGACGTCTAGAATGGTGAACTCTTTCCATAAGT 44819

Qy       1856 TTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTAT 1915
              ||||  |||||||| |   |||||  |      ||||||  ||||     || |||  ||
Db      44820 TTTTTTATTTACTTTGCTTAGATCTGT-----GAGAGGAAACACT----GTGACAGCCAT 44870
```

```
Qy       1916 AGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATC 1975
Db      44871 AACCTTACTAAATGTGTTTCCCAAATAATAAGATTTGAAAGT--CAAATTGCTCCTTATC 44928

Qy       1976 CATTTTCTGCAAAATAGATGTTGTG 2000
Db      44929 CATGGGCTGCAGAATGGATGTTGTG 44953


RESULT 17
US-09-949-016-14772
; Sequence 14772, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14772
;   LENGTH: 92681
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(92681)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14772

  Query Match             22.3%;  Score 446;  DB 3;  Length 92681;
  Best Local Similarity   72.9%;  Pred. No. 3.3e-82;
  Matches  757;  Conservative    0;  Mismatches  225;  Indels   56;  Gaps   12;

Qy        955 TTTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCA 1014
Db      16453 TTTGATGGAAAAGTTTGAGAAGTATAGGCATACCTCAGAGATATAGCAGGCCCAGCTCCG 16512

Qy       1015 TACCACCACAATAATACAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAA 1074
Db      16513 GGCCACCACAATAAAGCA--------------GATATCACAATAAAGTGAGTCACATGA 16557

Qy       1075 GTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAG 1134
Db      16558 ACTTTTTGGTTTCCCGGTGCATATAAAAGTTATGTTTACACTATACTGTGTTCTATTAAG 16617

Qy       1135 TGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAATTTTAAAATGCTTTATTACTA 1194
Db      16618 TGTGCAATAGCATTATGCCTAAACAAATTACATACCTTAATTAAAAATACTGTATT-GTT 16676

Qy       1195 AAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGT 1254
Db      16677 TAAAATGCTAACAATTATCTAAGCCTTCAGCAAATCATAATCTTTGTTGGTGAAGGATCT 16736

Qy       1255 C---------TTTTCTTATTGATGACTGATCGGGGGTCAGG-TGCTGAAGCTTAGGGTGGC 1305
Db      16737 CACCTTGATGTCGTTGGCTGCTGACTGATCAGGGTTGTGGTTGCTGAAGGTTAGGGTAAC 16796

Qy       1306 TGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT 1360
Db      16797 TGGAGCAATTTCTTAAAATAAGACAACGATGAAATTTGCCATGCTAATTGATTCTTCCTT 16856

Qy       1361 TCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAA 1420
Db      16857 TCATGAAAGATTTCTCTGTGGCATGTGATTCTGTTTGATAGCATTTTACTCATAGCAGAA 16916

Qy       1421 CTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTA 1479
Db      16917 CTTCTTTCAGACTTGGAGTCAGTCCTCTCAAACCCTACCACTGCCTTACCAACTAAGTTT 16976

Qy       1480 ATATAATATTCTGAATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGA 1539
Db      16977 ATGGAATACTCTAAGTCCTGTGTTGTCATTTCCACAATGTTCACAGCATCTTCACCAGGA 17036

Qy       1540 GTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTC 1599
Db      17037 GTAGATTCCATCTCAGGAAACCACTT----TCTTTGTTCATCCAGAAGAAATAACTCTAC 17092

Qy       1600 ATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTC 1659
Db      17093 ATCTGTTCAAATTTTAT--TGAAATTGCAGCAATTCAGTCTCATCTTCGGG-----CTTT 17145

Qy       1660 ACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAA 1719
Db      17146 ACTTCTAATTCCAGTTCTCTTGCTATTCCCACCACATCTGTAGTTGCTTCCTCCATTGAA 17205

Qy       1720 GTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTT 1779
Db      17206 GTCTCAAACCCCTCAAAGTCAGTCATCAGGGTTGGAATCAACTTCTTCCAAACTCCTATT 17265

Qy       1780 AATATTTATATTTTGACC---TCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATG 1836
Db      17266 AATGTTGATATTTTGACCTATTCCCATGAATTACAAATGTTCTTAATGTCATCTAGAATG 17325

Qy       1837 GTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATC 1896
Db      17326 GTGAATCATTTCCACAAGATTTTTAATTTACTTTACCCAG----ACCCATCAGCAGAATC 17381

Qy       1897 CACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAG 1956
Db      17382 ACAAATCTGTGGCAGCAATAGCATTATAAAATGTAGGTCT-------TAAGACTTGAAAT 17434

Qy       1957 TTGAAATTACTCCTTGAT 1974
Db      17435 TTGAAATTGCTCTTTAAT 17452


RESULT 18
US-09-949-016-12731/c
; Sequence 12731, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12731
;   LENGTH: 264206
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12731

  Query Match             22.2%;  Score 444.2;  DB 3;  Length 264206;
  Best Local Similarity   74.9%;  Pred. No. 9.7e-82;
  Matches  724;  Conservative    0;  Mismatches  208;  Indels   35;  Gaps   12;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
```

```
Db      28076 AAGTGAATATTGCAATAAAGCAAGTCACAAGAATATTTTGGTTTCCCAGTGCATATAAAA 28017

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC 1162

Db      28016 GCTACATTTACATTACACAATGGTCTATTAAGTGGGCAATAGTATTATGTTTAAAAA--- 27960

Qy       1163 ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTC 1222

Db      27959 ATACACTTAGTTAAAAACATTCTTTATTACTAAAAAATACTAACAGTCACCTGAGTCTTG 27900

Qy       1223 AGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGGT 1282

Db      27899 AGCAATTGGTAATCTTTTTGCTGGTCTTGCCTCATTGTTGATAGCTGCTCAGT--GGAGG 27842

Qy       1283 CAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA-----CAACAGTGAAGATT 1337

Db      27841 TCTTTGCTGAAGCTGGGGGTGTCTATGGCAATTTCTTAAAATAAGGCAACAATGAAGCTT 27782

Qy       1338 GCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTG 1397

Db      27781 CCCACATCAACTGACTCTT-CTTTCATGAATGATTTGTTTGTAGCATGCAATGCCACTTG 27723

Qy       1398 ATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTG 1456

Db      27722 ATAGCATTTTACCCACGCTAGAACTTCTTTCAAAATTGGAGTCAATCATCTCCAACTTTA 27663

Qy       1457 CTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAA 1516

Db      27662 ATGCTGCTTTATCAACTAAGTTGATGGAATATTCTAAATC---TTTTGTCATTTTGACAA 27606

Qy       1517 TTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGC 1576

Db      27605 TGTTCACAGCATTTCCACCAGGAATAGCTTCCATCTCAAGAAACCACTTTCTTTGCTT-A 27547

Qy       1577 TCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACA 1636

Db      27546 TCTTTAAGAAAAAGCAACACCTCATCCATTCCAGTTTTATCAGGAGATTGCAGCAATTCA 27487

Qy       1637 GTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACAT 1696

Db      27486 GTCATATCTTCAGG-----CTCCACTTCTAATTCTAGTTCTCTTGCTATTTCTACCACAT 27432

Qy       1697 CTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAA 1756

Db      27431 CTGCACTTACTTCCTCCACTAAAGTCTTGAAGCCTTCCAAGTCATCCATGAGAGCTGGAA 27372

Qy       1757 TCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATCATGAA 1813

Db      27371 TCAACTTCTTCCAAACTTCTGCTAATGTTGATGTTTTGACATCCTCCCATGAATCATGAA 27312

Qy       1814 TGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTC 1873

Db      27311 TGTTCTTAATGGTATCTCAAATGATGTATCCTTTCCAGAAGGTTTTCAACTCACTTTGTC 27252

Qy       1874 CAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATT 1933

Db      27251 CAGATCCA-----CCAGAGGAATCACT----ATGGCAGGTATAGCTTTATGAAATGTATG 27201

Qy       1934 TCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGA 1993

Db      27200 TCTTAAATAAT--GACTTGAATTTTGAAATTTCTCCTTGATCCAAGGACTGCTGAAAGGA 27143

Qy       1994 TGTTGTG 2000

Db      27142 TGTTGTG 27136


RESULT 19
US-09-949-016-13249/c
; Sequence 13249, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13249
;   LENGTH: 264304
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13249

  Query Match             22.2%;  Score 444.2;  DB 3;  Length 264304;
  Best Local Similarity   74.9%;  Pred. No. 9.7e-82;
  Matches  724;  Conservative    0;  Mismatches  208;  Indels   35;  Gaps   12;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102

Db      28076 AAGTGAATATTGCAATAAAGCAAGTCACAAGAATATTTTGGTTTCCCAGTGCATATAAAA 28017

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC 1162

Db      28016 GCTACATTTACATTACACAATGGTCTATTAAGTGGGCAATAGTATTATGTTTAAAAA--- 27960

Qy       1163 ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTC 1222

Db      27959 ATACACTTAGTTAAAAACATTCTTTATTACTAAAAAATACTAACAGTCACCTGAGTCTTG 27900

Qy       1223 AGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGATCGGGGGT 1282

Db      27899 AGCAATTGGTAATCTTTTTGCTGGTCTTGCCTCATTGTTGATAGCTGCTCAGT--GGAGG 27842

Qy       1283 CAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA-----CAACAGTGAAGATT 1337

Db      27841 TCTTTGCTGAAGCTGGGGGTGTCTATGGCAATTTCTTAAAATAAGGCAACAATGAAGCTT 27782

Qy       1338 GCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTG 1397

Db      27781 CCCACATCAACTGACTCTT-CTTTCATGAATGATTTGTTTGTAGCATGCAATGCCACTTG 27723

Qy       1398 ATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTG 1456

Db      27722 ATAGCATTTTACCCACGCTAGAACTTCTTTCAAAATTGGAGTCAATCATCTCCAACTTTA 27663

Qy       1457 CTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAA 1516

Db      27662 ATGCTGCTTTATCAACTAAGTTGATGGAATATTCTAAATC---TTTTGTCATTTTGACAA 27606

Qy       1517 TTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGC 1576

Db      27605 TGTTCACAGCATTTCCACCAGGAATAGCTTCCATCTCAAGAAACCACTTTCTTTGCTT-A 27547

Qy       1577 TCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACA 1636

Db      27546 TCTTTAAGAAAAAGCAACACCTCATCCATTCCAGTTTTATCAGGAGATTGCAGCAATTCA 27487

Qy       1637 GTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACAT 1696

Db      27486 GTCATATCTTCAGG-----CTCCACTTCTAATTCTAGTTCTCTTGCTATTTCTACCACAT 27432

Qy       1697 CTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAA 1756

Db      27431 CTGCACTTACTTCCTCCACTAAAGTCTTGAAGCCTTCCAAGTCATCCATGAGAGCTGGAA 27372

Qy       1757 TCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATCATGAA 1813
```

```
Db      27371 TCAACTTCTTCCAAACTTCTGCTAATGTTGATGTTTTGACATCCTCCCATGAATCATGAA 27312

Qy       1814 TGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTC 1873
              |||||||||||| | ||  |||| || |||||||||| |||||||||| | |||| |||
Db      27311 TGTTCTTAATGGTATCTCAAATGATGTATCCTTTCCAGAAGGTTTTCAACTCACTTTGTC 27252

Qy       1874 CAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATT 1933
              ||||||||     ||||||||  ||||    ||| ||| |||||| ||||| |||||||
Db      27251 CAGATCCA-----CCAGAGGAATCACT----ATGGCAGGTATAGCTTTATGAAATGTATG 27201

Qy       1934 TCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGA 1993
              |||| ||||||  | ||||||  |||||||| ||||||||||||     ||||  ||  ||
Db      27200 TCTTAAATAAT--GACTTGAATTTTGAAATTTCTCCTTGATCCAAGGACTGCTGAAAGGA 27143

Qy       1994 TGTTGTG 2000
              |||||||
Db      27142 TGTTGTG 27136


RESULT 20
US-09-949-016-14125/c
; Sequence 14125, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14125
;  LENGTH: 177797
;  TYPE: DNA
;  ORGANISM: Human
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(177797)
;  OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14125

  Query Match             22.1%;  Score 443;  DB 3;  Length 177797;
  Best Local Similarity   73.0%;  Pred. No. 1.6e-81;
  Matches  725;  Conservative    0;  Mismatches  215;  Indels   53;  Gaps   10;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
              ||||   |||||||||||||| || |||||  | ||| |  | |||||||||| ||||||
Db      11287 AAGTTATTATCACAATAAAGCGACTCACATGAATCTGTCAGTTTCCCAGTGCAAATAAAA 11228

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC 1162
              ||  || ||| |||| ||||| |||| |||||| |||| ||| | ||||  |   |||
Db      11227 GTCATGTTTACACTATACTGTCGTCTATTAAGTTTGCAGTAGCGCTATGCTT--TAAAGT 11170

Qy       1163 ACATACCTTAATTTTAAAATGCTTTATTAC-TAAAAAATGCTAACAATCATTTGAGCATT 1221
              |||||||||||||||||  |||||| | | |||||||||| ||||||  |||| ||
Db      11169 ACATACCTTAATTTTAAAACACTTTATTGCTTTAAAAATGCTAATAATCATCGGAGCCTT 11110

Qy       1222 CAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGA---------- 1270
              | ||||||  ||| |||||||||||| || ||||||  |||  || |||
Db      11109 CTGTGAGTCATAACCTTTTTGCTGGTAGAGGGTCTTGCCTTGATGTTGATGACTGTCAAC 11050

Qy       1271 CTGATCGGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACA 1325
              |     ||  |   | | |||| | || |||||||||||||| |||||||||     |||
Db      11049 CAATCAGGATGGGGGTTCCTGAGGGTTGGGGTGGCTGTGGCAATTTCTTAAAATAAGACA 10990

Qy       1326 ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTG 1385
              |||||||||  ||||   | |  |||| |||||||| |  ||| |||||||| ||||  |
Db      10989 ACAGTGAAGTCTGCAGCTTTACCTGACCCTTCCTTTTACTAAATATTTCTCTGTAGTAAG 10930

Qy       1386 TGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCT 1445
              |||| || ||||||||||||| |  ||   ||||||| |||| |||| ||  ||| ||||
Db      10929 TGATACTGTTTGATAGCATTTGACCCATGCTAGAACTGCTTTCAAAA-TGAGTCAGTCCT 10871

Qy       1446 CTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGT 1505
              || ||  ||||||     || |  |  ||| || |||   |||||| ||||||||||||
Db      10870 CTTAAGTCCTGCTAATGCTTTATCAATTAAATTTATAGCGTATTCTAAATCCATTGTTGT 10811

Qy       1506 CATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGAT 1565
              |||||| |||||  |||||||  ||| ||||||||||| |||| ||||||||      | |
Db      10810 CATTTC-ACAATGTTCACAGCATCTACACCAGGAGTCGATTTCATCTCATGAAAGCACTT 10752

Qy       1566 GGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATT 1625
                |||||||||||  |||||| || |  ||||||||  |||||||||  | |||||||||
Db      10751 ----TCTTTGCTCATTCATAAGCAGCAGCTCCTCATCCATTCAAGTTTGAACATGAGATT 10696

Qy       1626 GCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGT 1685
              ||||||  |   ||||| ||||||||       | ||||| ||| || ||||||||||||  |
Db      10695 GCAGCAGTTTGGTCATATCTTCAGG-----ATCCACTTCTAACTCGAGTTCTCTTGCTAT 10641

Qy       1686 TTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCAT 1745
              ||| |||| |||||| ||  |||||||| |||| | ||| ||| ||| ||||||||||
Db      10640 TTCCACCATATCTGTAGTGACTTCCTCCACTGAAATTTTGTACCCCTCAAAGTCATCCAT 10581

Qy       1746 GAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---------- 1795
                |||||  ||| | ||| | ||| |   ||   || |  |  |||| || |
Db      10580 AAGGGTGGGATTTGACATTTTCAAGTATCTAATTTAATTAAATATCTTTAATGCTGATAT 10521

Qy       1796 --------CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
                     ||||||||| ||||||||||||  |||||||| || ||||||  ||  ||||
Db      10520 TTTGACCTCCTCCCATGTATCATGAATGTTTCTAATGGCATCTAGAATGGCAAACTCTTT 10461

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
              ||| || |  |||  |||||||| | ||||||||||     |||||||  ||   || |||
Db      10460 CCAGAAAGCTTTTTATTTACTTTGCCCAGATCCAT-----CAGAGGAATCATGATCTATG 10406

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
                |||| |||||||||   |||||||||||| |||||||||  |||||||||  |||| |||
Db      10405 GCAGTAATAGCCTTAAAAAATGTATTTCTTAAATAATAAGACTTGAAAGTCCAAATGACT 10346

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              ||||||||||   || || ||| |||||||||
Db      10345 CCTTGATCCATGGGCTACAGAATGGATGTTGTG 10313


RESULT 21
US-09-949-016-11842/c
; Sequence 11842, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11842
;  LENGTH: 227979
;  TYPE: DNA
;  ORGANISM: Human
;  FEATURE:
;  NAME/KEY: misc_feature
```

```
;   LOCATION: (1)...(227979)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11842

  Query Match             22.1%;  Score 443;  DB 3;  Length 227979;
  Best Local Similarity   73.0%;  Pred. No. 1.7e-81;
  Matches  725;  Conservative    0;  Mismatches  215;  Indels   53;  Gaps   10;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
             ||||   |||||||||||||| || |||||  | ||| |  | |||||||||| ||||||
Db      13469 AAGTTATTATCACAATAAAGCGACTCACATGAATCTGTCAGTTTCCCAGTGCAAATAAAA 13410

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC 1162
             ||  || ||| |||| ||||| |||| |||||| |||| ||| | ||||   |    |||
Db      13409 GTCATGTTTACACTATACTGTCGTCTATTAAGTTTGCAGTAGCGCTATGCTT--TAAAGT 13352

Qy       1163 ACATACCTTAATTTTAAAATGCTTTATTAC-TAAAAAATGCTAACAATCATTTGAGCATT 1221
             ||||||||||||||||||  ||||||| | | |||||||||| ||||||  |||| ||
Db      13351 ACATACCTTAATTTTAAAACACTTTATTGCTTTAAAAATGCTAATAATCATCGGAGCCTT 13292

Qy       1222 CAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGA---------- 1270
             | ||||||  ||| |||||||||||| || ||||||  |||  || |||
Db      13291 CTGTGAGTCATAACCTTTTTGCTGGTAGAGGGTCTTGCCTTGATGTTGATGACTGTCAAC 13232

Qy       1271 CTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACA 1325
             |      ||  |   | | |||| | || ||||||||||||| |||||||||     |||
Db      13231 CAATCAGGATGGGGGTTCCTGAGGGTTGGGGTGGCTGTGGCAATTTCTTAAAATAAGACA 13172

Qy       1326 ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTTG 1385
             |||||||||  ||||   | |  |||| |||||||| |  ||| |||||||| ||||  |
Db      13171 ACAGTGAAGTCTGCAGCTTTACCTGACCCTTCCTTTTACTAAATATTTCTCTGTAGTAAG 13112

Qy       1386 TGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCT 1445
             ||||  || ||||||||||||||  |  ||   ||||||| |||| |||| ||  ||| ||||
Db      13111 TGATACTGTTTGATAGCATTTGACCCATGCTAGAACTGCTTTCAAAA-TGAGTCAGTCCT 13053

Qy       1446 CTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGT 1505
             || ||  ||||||     || |  |  ||| || |||   ||||||  |||||||||||||
Db      13052 CTTAAGTCCTGCTAATGCTTTATCAATTAAATTTATAGCGTATTCTAAATCCATTGTTGT 12993

Qy       1506 CATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGAT 1565
             ||||||  |||||  |||||||   ||| ||||||||| |||| ||||||||      | |
Db      12992 CATTTC-ACAATGTTCACAGCATCTACACCAGGAGTCGATTTCATCTCATGAAAGCACTT 12934

Qy       1566 GGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATT 1625
              ||||||||||| |||||| || |  ||||||||  |||||||| | |||||||||
Db      12933 ----TCTTTGCTCATTCATAAGCAGCAGCTCCTCATCCATTCAAGTTTGAACATGAGATT 12878

Qy       1626 GCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGT 1685
             ||||||  |   ||||| ||||||||       | ||||| ||| || ||||||||||| |
Db      12877 GCAGCAGTTTGGTCATATCTTCAGG-----ATCCACTTCTAACTCGAGTTCTCTTGCTAT 12823

Qy       1686 TTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCAT 1745
             |||  |||| |||||| ||  |||||||| |||| | ||| ||| ||| ||||||||||
Db      12822 TTCCACCATATCTGTAGTGACTTCCTCCACTGAAATTTTGTACCCCTCAAAGTCATCCAT 12763

Qy       1746 GAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---------- 1795
              ||||| ||| | ||| | ||| |   ||   || |  |  |||| || |
Db      12762 AAGGGTGGGATTTGACATTTTCAAGTATCTAATTTAATTAAATATCTTTAATGCTGATAT 12703

Qy       1796 --------CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTT 1847
                     ||||||||| ||||||||||||  |||||||| || ||||||  ||  ||||
Db      12702 TTTGACCTCCTCCCATGTATCATGAATGTTTCTAATGGCATCTAGAATGGCAAACTCTTT 12643

Qy       1848 CCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATG 1907
             ||| || | |||  |||||||| | ||||||||||     |||||||  ||   || |||
Db      12642 CCAGAAAGCTTTTTATTTACTTTGCCCAGATCCAT-----CAGAGGAATCATGATCTATG 12588

Qy       1908 CCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACT 1967
              |||| |||||||||   |||||||||||| |||||||||  |||||||||  |||| |||
Db      12587 GCAGTAATAGCCTTAAAAAATGTATTTCTTAAATAATAAGACTTGAAAGTCCAAATGACT 12528

Qy       1968 CCTTGATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||||||||||   || || ||| |||||||||
Db      12527 CCTTGATCCATGGGCTACAGAATGGATGTTGTG 12495


RESULT 22
US-09-497-855A-38
; Sequence 38, Application US/09497855A
; Patent No. 6605432
; GENERAL INFORMATION:
;  APPLICANT: Huang, Tim
;  TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION
;  FILE REFERENCE: UMO1523
;  CURRENT APPLICATION NUMBER: US/09/497,855A
;  CURRENT FILING DATE:  2000-02-04
;  PRIOR APPLICATION NUMBER: 60/120,592
;  PRIOR FILING DATE: 1999-02-18
;  PRIOR APPLICATION NUMBER: 60/118,760
;  PRIOR FILING DATE: 1999-02-05
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38
;   LENGTH: 128779
;   TYPE: DNA
;   ORGANISM: Homo sapiens;
US-09-497-855A-38

  Query Match             22.0%;  Score 439.8;  DB 3;  Length 128779;
  Best Local Similarity   72.7%;  Pred. No. 6.7e-81;
  Matches  747;  Conservative    0;  Mismatches  232;  Indels   48;  Gaps   12;

Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
              |||||||||| ||||| ||| ||  | |||||||| | |||||  ||| ||  ||||||
Db      42749 TACAGGCATACATTGGACATATTGCAGATTTGGTTCTAGACCACTGCAAAAAAGCAAATA 42808

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096
              |                |||||||||||||| |||||| | |||||  ||| |||||||||
Db      42809 TTG---------------CAATAAAGTGAGTTACACAATTTTTTTGCCTTTCCAGTGCAT 42853

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156
              ||||||||| || ||||||||||     ||||| |||||||||| |||  |        ||
Db      42854 ATAAAAGTTATGTTTATACTACATGACAGTCTATTAAGTGTGCAGTAGCAT--GTCTAAA 42911

Qy       1157 AAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGA 1216
              |||  || ||||||||||  || || |||||||| ||||||||||  |  | || |
Db      42912 AAATGCATATACCTTAATTAAAAGATACTTTATTGCTAAAAAATGC-AGGACTCTTCACT 42970

Qy       1217 GCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGA-T 1275
              |  |      |||          |          | | | || | | ||| |||| |
Db      42971 GAGTCATAATCTTTGAGGCTGCAGAGTCTTACCTTAATGTCTGCTGACTCATGGCTGATT 43030

Qy       1276 CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGT 1330
              ||| |  |  |||||||||  |  |||||||||||||| |||||||||     |||||| |
Db      43031 AGGGTGATAACTGCTGAAGGCTCAGGTGGCTGTGGCAATTTCTTAAAACATGACAACACT 43090

Qy       1331 GAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATG 1390
              |||| ||||   |||||||||| ||||||||||||||||   |||||||| | || |||||||
Db      43091 GAAGTTTGCTGCATCAGTTGATTCTTCCTTTCATGAACTATTTCTCTGTGGTATGTGATG 43150

Qy       1391 CTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAA 1450
              || ||| |||||||||||  |||||  |        |  ||  | |  ||||||||||||
Db      43151 CTGTTTAATAGCATTTTACCCACAGAA-----CTCTCAAAGTTGGAGTCAATCCTCTCAA 43205

Qy       1451 ACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTT 1510
              || ||||       || |  |  ||||| || | | ||||| ||| | ||| ||| ||||
Db      43206 ACTCTGCCACTGCTTTATCAGTCAAGTTTATGTGACATTCTAAATTCTTTGCTGTTATTT 43265

Qy       1511 CAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAAT 1570
              |||||  |||||||  |||||  |||||||||||||| ||||  |        |  |    |
Db      43266 CAACAA-TTTCACAGCATCTTCCCCAGGAGTAGATTTCATCCTAAGAAACCACTT----T 43320

Qy       1571 CTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGC 1630
              |||| |||||||||||||| || | ||||||  |||||  ||||||||| ||| ||||||
Db      43321 CTTTACTCATCCATAAGAAGCAACTTCTCATCCATTCAAATTTTATCATAAGAATGCAGC 43380
```

```
Qy       1631 AATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTA 1690
Db      43381 AATTCTGTCACATCTTCAGG-----CTACACTTCTAATTCTAGTTCTCTTGCTATTTCCA 43435

Qy       1691 CCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGG 1750
Db      43436 CTATATCTACAGTTATTTCCTCCACTGAAGTCTTGAACCTGTGAAAGTCATCCATGAGGA 43495

Qy       1751 TTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAAT 1807
Db      43496 TGGGAATCAACTTCTTCCAAACTCCTGTTCATGTTGATATTTTGATCTCCTCCCATGAAT 43555

Qy       1808 CATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTAC 1867
Db      43556 CACAAATGTTCTTAATTGCATCTAGAAT-GTGAATCCTCTCCAGAAGGTTTTCAATGTAC 43614

Qy       1868 TTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAA 1927
Db      43615 TTTGCTCAGATCCAT-----CAGAGGAATCACTATCTATGGCAGCTATACCCTTACAAAA 43669

Qy       1928 TGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAA 1987
Db      43670 TGTATTTCTTAAATAATAAGACTTGAAAGTTGAAATTACTCCTTGATCTGTGGGCTGTGG 43729

Qy       1988 AATAGAT 1994
Db      43730 AATGGTT 43736
```

RESULT 23
US-09-949-016-14083
; Sequence 14083, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14083
;   LENGTH: 199471
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(199471)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14083

  Query Match   21.9%; Score 438.4; DB 3; Length 199471;
  Best Local Similarity   76.3%; Pred. No. 1.4e-80;
  Matches 734; Conservative   0; Mismatches 181; Indels   47; Gaps   14;

```
Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
Db      23871 AAGCAAGTATTGCAATAAAGTGAGTCACACAATTTTTTTTATTTCTCAGTGCATGTAAAA 23930

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC 1162
Db      23931 GTTAGG-TTACTATATACTGTAGTCTATTAAGTATGCAATAGCATAATGTCTTAAAAACC 23989

Qy       1163 A------CATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGA 1216
Db      23990 AAATGTGCATACTTTAATTTTAAAATACGTTATTGCTAAAAAA---AAATAATAATCTGA 24046

Qy       1217 GCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGAT-------- 1268
Db      24047 GCCTTCAGTGAGTTGTAATCTTTGTGCTAATGAAGGGTCTTGTCTTGATGTTGTTGGCTG 24106

Qy       1269 --GACTGATCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA--- 1323
Db      24107 CGGACTGATCAGGATGATGGTTGCTGAAGTTGGGGTGGCTATGGCAATTGCTTAAAATAA 24166

Qy       1324 --CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAG 1381
Db      24167 GTCAACAGTGAAGTTTACCACATTGATTGACTCTTCCTTTAATCAAATATTTCTGTGCAG 24226

Qy       1382 TGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCA 1440
Db      24227 CATGTTGTGCTGTTTGTTAGCATTTCACCCACAGTAGAACTTCTTGCAAAATTGGAGTCA 24286

Qy       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500
Db      24287 ATTCCCTCAAACCCTGCCACTGCTTTATCAACTAA-ATATTATAATATACTAAATCCTGT 24345

Qy       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560
Db      24346 GTTGTC-TTTCAACAATGTACACGGCATCTTCACCAGAAGTAGATTCTATCTCAAGAAAC 24404

Qy       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
Db      24405 CACTT----TCTTTATTCATCCATAAGAAGCAACTTCTCATCCATTCAAGTCTTATCATG 24460

Qy       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680
Db      24461 AGATTACAGCAATTCAGTCATATCTTCAGGCCCCAGTTC-----TAATTCTATTTCTCTT 24515

Qy       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCA 1740
Db      24516 GCTATTTCCGCCACATCTGCAGTTACTTCCTCCACTGAAGTCTTGAACCCCTCCAAGTCA 24575

Qy       1741 TCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTT-AATATTTATATTTTGAC--- 1796
Db      24576 TCCTTGACGACTGGAGTCAACTTCTTCCAAACTCCTGTTAAATGTTGATATTTTGACCTT 24635

Qy       1797 CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGT 1856
Db      24636 CTTCCATCAATCACAAGTGTTTCTAATGGCATCTAGAATGGTGAATACTTTCCAGAATGT 24695

Qy       1857 TTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATA 1916
Db      24696 TTTCAATGTACTTTGTCCAGATCCAT-----GAGAGGAATAACTATCTATG-CAGCTATA 24749

Qy       1917 GCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCC 1976
Db      24750 GCCTTGCAAAATGTATTTCTTAAATAACAAGACTTGAAAGTCTAAATTACTCCTTGATCC 24809

Qy       1977 AT 1978
Db      24810 AT 24811
```

RESULT 24
US-09-949-016-13969
; Sequence 13969, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768

```
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13969
;   LENGTH: 301828
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(301828)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13969

  Query Match            21.9%;  Score 437.4;  DB 3;  Length 301828;
  Best Local Similarity   69.9%;  Pred. No. 2.5e-80;
  Matches  797;  Conservative    0;  Mismatches  261;  Indels   83;  Gaps   12;

Qy        875 TTTTTATGCTATGTCTATAATACATTAGGTTTGTTGTTGTTTGTTTTTACCTTATTGCTG 934
              ||||| || |     | || |   |  |   |  |   ||  ||| | ||  |
Db     224737 TTTTTCTGTTCCTGTTGTAGCAGGGCATTTAGTTCACTTCCTGAGATTAATTGATGCCAA 224796

Qy        935 TTGGCTGGGGTCAGCAAACATTTTCTGTAAAGGGCTAGATAGTACAGGCATACCTTGGAG 994
              ||||     |       |   | ||  |||     |  |  | |||||| |||| |||
Db     224797 GTGGCATTTCTTGCTCTGCGAATCCTCAAAACAAAGATCTTTTGCAGGCACACCTCAGAG 224856

Qy        995 ATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATATGCAAGAAGTGGATATCA 1054
              ||| ||  || |  |||||| |||||||||| |                  ||||| |||||
Db     224857 ATATTGCAGGCTCAGTTCCAGACCACCACAACA--------------AAGTGAATATCG 224901

Qy       1055 CAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCTTATA 1114
              ||||||||  ||| |||||  | ||||  ||    ||||||||||||||||  | ||| |
Db     224902 CAATAAAGCAAGTAACACACATTTTTTTTCT----AGTGCATATAAAAGTTAGGTTTACA 224957

Qy       1115 CTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAACACATACCTTAAT 1174
              ||   |||||||| |||||||||                          |||||||| ||
Db     224958 CTCTTCTGTAGTCTATTAAGTGTGT----------------------ACATACCTTCAT 224994

Qy       1175 TTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTGAGTTGTAA 1234
              |  || |    ||     |  ||||||||||  |||||| ||||   |||||||||| |||
Db     224995 TAAAATAA---TTTACTCCTAAAAATGCTAGTAATCATCTGAGTCATCAGTGAGTTTTAA 225051

Qy       1235 TCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGA----------CTGATCGGGGGTCA 1284
              |||| || ||||| || ||||||  || | || ||          ||||||||||  |
Db     225052 TCTTCTTTCTGGTAGAGGGTCTTGCCTCAGTGTTGGTGGCTGCTGTCTGATCGGGGGGTG 225111

Qy       1285 GGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----ACAACAGTGAAGATTGC 1339
              |  ||||||| ||  |||||||| | || |||||||||     ||||||  ||||||||||
Db     225112 GTAGCTGAAGATTGTGGTGGCTGCGACAATTTCTTAAAATAAGACAACAATGAAGATTGC 225171

Qy       1340 AATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCTTTTTGAT 1399
                 |||   |||||||||||| |  |||||||||| |||  |||||||| || |||
Db     225172 TGCATCGACTGACTCTTCCTTTGACAAAAGATTTCTCTGTAGCATGTGATGCTGTTGGAT 225231

Qy       1400 AGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCTCAAACCCTGCT 1458
              |||||||||  || |||||||||||||| |||||||| |||||||| ||| || ||||||
Db     225232 AGCATTTTACCCAGAGTAGAACTTCTTTCAAAATTGGAGTCAATCCTTTCATACTCTGCT 225291

Qy       1459 CTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCAACAATT 1518
                 || |  | |||||||  || || ||||||  ||||| |||||||||||||||||||
Db     225292 GCTGCTTCATCAACTAAGTTTATGTACTATTCTAAATCCTTTGTTGTCATTTCAACAATG 225351

Qy       1519 TTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTC 1578
              ||||||  |||||||  |||||||| ||||||||||        |  |    ||||||||
Db     225352 CTCACAGAATCTTCACTAGGAGTAGCTTCCATCTCAAGAGACCACTT----TCTTTGCTA 225407

Qy       1579 ATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGT 1638
              | || ||||||| || ||||||| |||||||||| |||||||| |||||||||| ||||
Db     225408 ACCCGTAAGAAGCAACTCCTCATATGTTCAAGTTTGATCATGAGTTTGCAGCAATTCAGT 225467

Qy       1639 CATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCT 1698
              ||  | ||||||     || ||||| |||||| | ||||||| || |||| |||||||||
```

```
Db     225468 CACATGTTCAGG-----CTCCACTTCTAATTCTATTTCTCTTACTTTTTCCACCACATCT 225522

Qy       1699 GTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATC 1758
              |  ||| ||||||||| |||||||| || | ||| ||||||||||||||||||||||||
Db     225523 GCTGTTACTTCCTCCACTGAAGTCTTAAAGCCCTCAAAGTCATCCATGAGGGTTGGAATT 225582

Qy       1759 GACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTCCCATGAATCATGAATGTTC 1818
              |||||||||||| |||||||  ||||||   |  |   |||||||||||||| ||||||||
Db     225583 CACTTCTTCCAAACTCCTGTTGATATTTTCACCT---CCTCCCATGAATCACGAATGTTC 225639

Qy       1819 TTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGAT 1878
              ||| | ||| || |||||| || || ||| || ||| |||||||||| || | ||||||
Db     225640 TTATTAGCATCTAGAATGGCGATTCTTTTTCAGAAGTTTTTCAATTTATTTTGCCCAGAT 225699

Qy       1879 CCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTC 1938
              ||||          |   | |  | ||| |||| |||| |||||  ||| ||||||||
Db     225700 TCATC---------AAAGGAATCACTATGGCAGTGATAG-CTTATTAAATATATTTCTTA 225749

Qy       1939 AATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTG 1998
              | ||||||| |||||||||  |||||||  |||||||||||    || || ||| |  ||||
Db     225750 AGTAATAAGTCTTGAAAGTCAAAATTACCCCTTGATCCATGGGCTACAGAATGGTAGTTG 225809

Qy       1999 T 1999
              |
Db     225810 T 225810


RESULT 25
US-09-949-016-16073/c
; Sequence 16073, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16073
;   LENGTH: 385136
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(385136)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16073

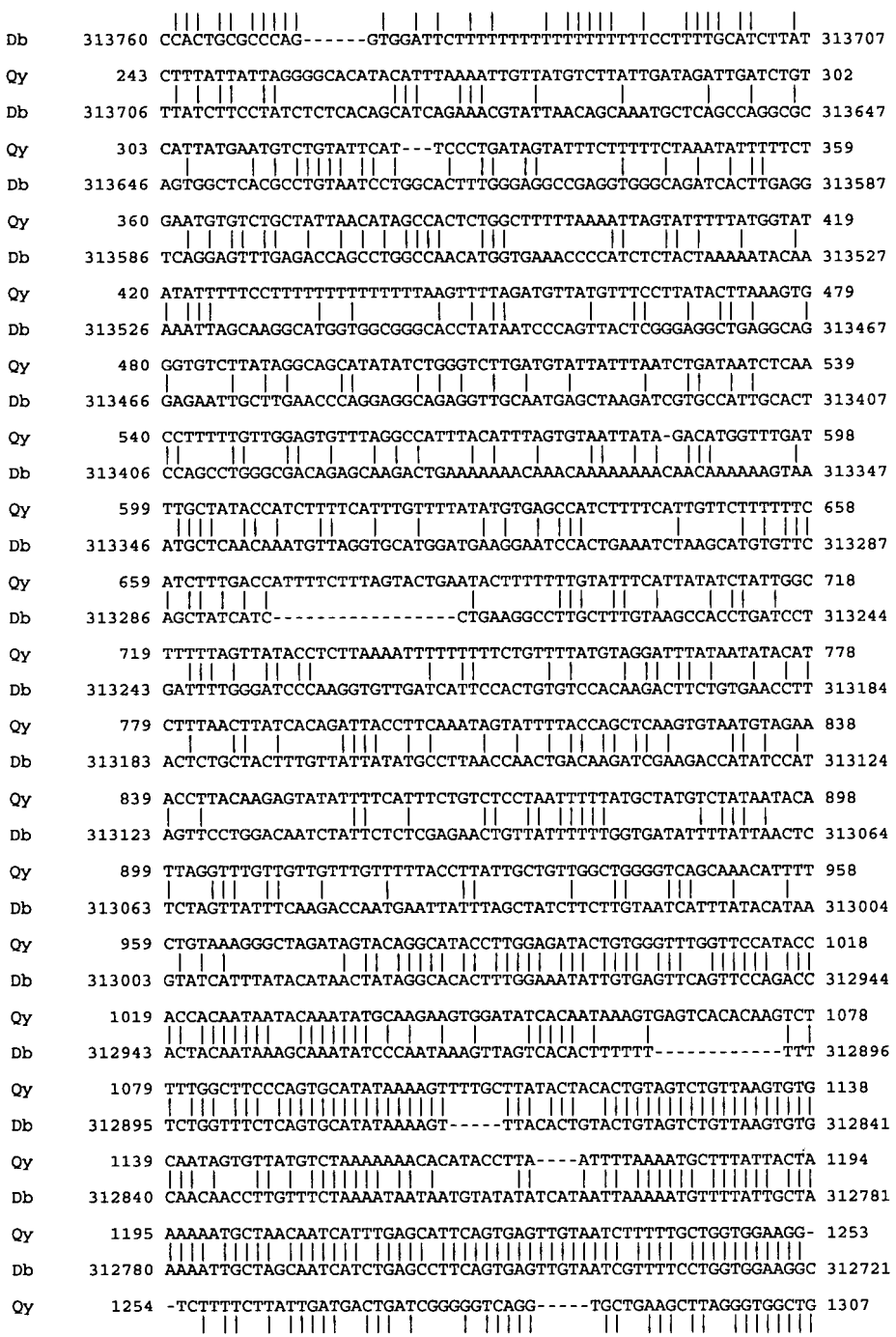
  Query Match            21.9%;  Score 437.4;  DB 3;  Length 385136;
  Best Local Similarity   57.6%;  Pred. No. 2.6e-80;
  Matches 1165;  Conservative    0;  Mismatches  746;  Indels  110;  Gaps   17;

Qy          3 ATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCCAGCCACCACGTCCAGC 62
              |||||||| ||||||||||  |||||||| |||||  |||      |||||||  || ||
Db     313940 ATTCTCCTCCCTCAGCCTCCTGAGTAGCTGGGGTTATAGGCATGCACCACCACGCCCGGC 313881

Qy         63 TAATTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTTTGCCAGGCTGGTATCAAACT 122
              |||||||    |||||||||||| |||||||| || |||| | |||||||||  |  ||||
Db     313880 TAATTTTGTATTTTTAGTAGAGATGGGGTTTCTCCATGTTGGTCAGGCTGGTCTTGAACT 313821

Qy        123 CCTGACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAATGCTAGGATTACAGGTGTGAG 182
              ||  |||||| |||| |  |  |||   |||||||||| |||| ||||||||||||||||
Db     313820 CCCAACCTCAGGTGACCCACTCGCCCTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAG 313761

Qy        183 TCACCGCACCCAGCCCTTCTTTCAGTTCTATCACCTCTTTTTGCTATATTTGTATGAGAG 242
```



RESULT 26
US-09-949-016-16433/c
; Sequence 16433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16433
; LENGTH: 96202
; TYPE: DNA

```
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(96202)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16433

  Query Match             21.8%;  Score 435.8;  DB 3;  Length 96202;
  Best Local Similarity   72.6%;  Pred. No. 4.2e-80;
  Matches  759;  Conservative    0;  Mismatches  232;  Indels   54;  Gaps   13;

Qy        978 ACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATAT 1037
              ||||||||| |||||||||| |  ||||||||||||| |||||  ||||
Db      89881 ACAGGCATATCTTGGAGATATAGCCGGTTTGGTTCCAGACCACGGCAAT---------- 89833

Qy       1038 GCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATA 1097
                ||||   ||||  ||| |||||||  |||||    ||||||  | |||||||| ||
Db      89832 ----GAAGGAAATATTGCAACAAAGTGAATCACATGCATCTTTTTGTTTCCCAGTCCACG 89777

Qy       1098 TAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAA 1157
              || ||||| |  ||| |||| |||  | | | |||||||||||| ||  |||| |  |||
Db      89776 TAGAAGTTATATTTACACTATACTAGAATGTATTAAGTGTGCAACAGCATTAT-TTAAAA 89718

Qy       1158 AAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAG 1217
              |    ||| ||||||| |  |||| |||||||| ||| |||||| ||||||||||  ||||
Db      89717 AGTGAACACACCTTAACTAAAAAACACTTTATTGCTAGAAAATGGTAACAATCATCTGAG 89658

Qy       1218 CATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-- 1275
                |||  |||||   |||||||||||||| |||| ||| || ||| | || |||  | |
Db      89657 TCTTCCCTGAGTCAAAATCTTTTTGCTGTTGGAGGGTGTTGTCTAAATGTTGATGGCTGC 89598

Qy       1276 ---------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA--- 1323
                      ||| |   | |||||||| || |||||||| || |||||||| ||||
Db      89597 TGATTGGCCAGGGTGGTGGTTGCTGAAGGTTGGGGTGGCTATGACAGTTTCTCAAAATAA 89538

Qy       1324 --CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAG 1381
                |||||  |||||  ||||   |||||||||||||   |  || |    | | |  | |
Db      89537 GGCAACAATGAAGTTTGCTGCATCAGTTGACTCTTTTGTGAATAATTTTTCTGTAGCATG 89478

Qy       1382 TGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCA 1440
                ||  |||||| ||| |||   |||| ||||||||||| | |||||| |||  ||| 
Db      89477 CAATGCTAGCAGTTTGATGGCAATTTACCCACAATAGAACTTCTTTCATAATTGGAGTCA 89418

Qy       1441 ATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATT 1500
              |||||||  ||  |||||     | |  | ||||||| || |||| | |  ||||| ||
Db      89417 ATCCTCTTAAGTCCTGCTGCTGCATTATCAACTAAGTTTATGTAATCTCCGAAATCCTTT 89358

Qy       1501 GTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCT 1560
              ||||||| ||||| ||| |||| ||  |||||| || |||||||||| ||||
Db      89357 GTTGTCACTTCAATAATCTTCATAGCATCTTCACTAGAAGTAGATTCCAACTCAAGAAAC 89298

Qy       1561 GAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATG 1620
                |  |    |||||||||||||||||  || |||||||| ||||||||||  || |
Db      89297 CACTT----TCTTTGCTCATCCATAAGAAACAACTCCTCATCT-TTCAAGTTTTACCACG 89243

Qy       1621 AGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTT 1680
              | |   |  |||| ||||||||||||||||     ||| |||| |||||  |||||||||
Db      89242 AAACCTCTTCAATTCAGTCATGTCTTCAGG-----CTTTACTTCTAATTTTAGTTCTCTT 89188

Qy       1681 GCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAAC-CTCTCCAAGTC 1739
              ||| ||| |  || |||||| |||  |||||||| |||||| |||||| | ||| |||||
Db      89187 GCTCTTT-TCACATATCTGTAGTTATTTCCTCCACTGAAGTTTTGAACACCCTCAAAGTC 89129

Qy       1740 ATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC--- 1796
              || || ||||| || |||  |||||||||||| || |||||||  ||  |||||| ||
Db      89128 ATGCAAGAGGGATGAAATAAACTTCTTCCAAACTCTTGTTAATGTTGCTATTTTCACCTT 89069

Qy       1797 CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGT 1856
              |||||||||||| |||||||||||||||| || ||| |||||||||||| | |||||
Db      89068 CTCCCATGAATCACGAATGTTCTTAATGGCATCTAGAAAGGTGAATCCTTTCTAGAAGGT 89009

Qy       1857 TTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATA 1916
              ||||||||||||   ||||||||||      |||||||  |||| || ||| |||  | |
Db      89008 TTTCAATTTACTTTACCCAGATCCAT-----CAGAGGAAACACTATCTATGGCAGCCACA 88954

Qy       1917 GCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTG-AAATTACTCCTTGATC 1975
              || |||   ||| |||||||| |||||||| | ||||||| | |||||||| ||||||
Db      88953 GCTTTACAAAATATATTTCTTAAATAATAAGACATGAAAGTGGAAAATTACTCTTTGATC 88894

Qy       1976 CATTTTCTGCAAAATAGATGTTGTG 2000
                ||   || || |||  ||||||||
Db      88893 TATGAACTACAGAATGAATGTTGTG 88869


RESULT 27
US-09-949-016-13319
; Sequence 13319, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13319
;   LENGTH: 103750
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13319

  Query Match             21.7%;  Score 434;  DB 3;  Length 103750;
  Best Local Similarity   72.7%;  Pred. No. 9.9e-80;
  Matches  759;  Conservative    0;  Mismatches  225;  Indels   60;  Gaps   13;

Qy        976 GTACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAAT 1035
              |||||| || ||  |||||||||| |||||| ||||||| |||||   |||||  |||
Db      43857 GTACAGACACACTGTGGAGATACTATGGGTTCGGTTCCAGACCACTCTAATAAAGCAA-- 43914

Qy       1036 ATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCA 1095
                          |||||||||||||||||| || ||  |   |||||| ||| |  | ||
Db      43915 -------------ATATCACAATAAAGTGATTCGCATGAACTTTTTGGTTTCTCCATACA 43961

Qy       1096 TATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTA 1155
              |||| | ||| || |||||| | ||||||||||   |||| ||||||||  ||||||  |
Db      43962 TATACATGTTATGTTTATACCATACTGTAGTCTAAGAAGTATGCAATAGGCTTATGTTCA 44021

Qy       1156 AAAAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTG 1215
              ||| |   |     | | |||||| | |||||||  | || | ||||||  ||||| ||
Db      44022 AAATACATCCATATCTTAATTTAAAGTACTTTATTTAT-AATAGTGCTAATGATCATCTG 44080

Qy       1216 AGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCT---TATTGATGACT 1272
              | | |||||  ||| ||||||||||  |||| ||| |||||| |||   |||||||| ||
Db      44081 ATCCTTCAGCAAGTGGTAATCTTTTCACTGGCGGAGGGTCTTGTCTCTATATTGATGGCT 44140

Qy       1273 GATCGGGGGTCAGG--------TGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAAC 1324
              |       | |||||        |||||||| || || || |||||||  || |||||||
Db      44141 GCCGACTGATCAGGATGGTGGCTGCTGAAGGTTGGGTTGACTGTGGCCATTCCTTAAAAT 44200

Qy       1325 A-----ACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCT 1379
              |     ||| ||||| |||| | || | ||| |||||||| ||||||||||||||||| |
Db      44201 AAGATGACAATGAAGTTTGCCACATTAATTGCCTCTTCCTCTCATGAAAGATTTCTCTGT 44260

Qy       1380 AGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATC 1439
              ||  |||||||  || |       ||||||  |||||||||||  ||  | | | |  |
Db      44261 AGCATGTGATG--TTGTTGATATATTTTACCCACAGTAGAACT--TTCAAGATTAGAGAC 44316

Qy       1440 AATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCAT 1499
```

```
Db      44317 CAACCTCTCAAACCCATTGCTGCTTTAATGA-CAAAGTTTTTGTAGTATCCTAAATCTTT 44375

Qy       1500 TGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCC 1559

Db      44376 TGTTGTTATTTCAATAATGATCACAGCATCTTCACAAGGATTTGATC---------TTCA 44426

Qy       1560 TGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCAT 1619

Db      44427 GAAACCACATTCTTTGCTTGTCCATTGGAAGCAACTCCTTATCCATTCAAGTTTGATTAT 44486

Qy       1620 GAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCT 1679

Db      44487 GAGATGGCAGCAGTTCAGTCACATTTTCAAG-----TTCCACTTCTAATTCTAGTTCTCT 44541

Qy       1680 TGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTC 1739

Db      44542 TGCTATTTCCACTGCATCTGCAGTTGCTTCCTCCACTGAAGTCTTGAACCCCTCAAAGTC 44601

Qy       1740 ATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTG---AC 1796

Db      44602 ATCCATGAGAGTTGAAATCAACTTCTTCCAAACTTCTATTAATGTTGATATTTTGTACTT 44661

Qy       1797 CTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGT 1856

Db      44662 CTCTCATGAATCACTAATCTTCTTAATGGCATGTAGAGTGGTGAATCTTTTTCAGAAGGT 44721

Qy       1857 TTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATA 1916

Db      44722 TTTCAATTTACTTTGCCCAGATTCAT-----CAGAGGAATCACCATCTATGACAGGTATA 44776

Qy       1917 GCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCC 1976

Db      44777 GCCCTATGAAATGTATTTCTTAAATAATAAGCCTTGAAAGTAAAAATT-TTCCTTGATCC 44835

Qy       1977 ATTTTCTGCAAAATAGATGTTGTG 2000

Db      44836 ATGGGCTGCAGAATGGATGTTGAG 44859


RESULT 28
US-09-949-016-13214
; Sequence 13214, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13214
;   LENGTH: 113701
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13214

  Query Match             21.7%;  Score 433.6;  DB 3;  Length 113701;
  Best Local Similarity   72.5%;  Pred. No. 1.2e-79;
  Matches  760;  Conservative    0;  Mismatches  229;  Indels   59;  Gaps   13;

Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036

Db       6984 TACAGGCATACCTTGGAGATAGTGTGGGTTCAGTTCCAGATCACCCCA------------ 7031

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096

Db       7032 ---AAGAAGTGAATATTGGAATACAGCGAGTCACACACGTTTTTTGGTTTTCTAGTGCAT 7088

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156

Db       7089 ATAAAAGTTATGTTTATACTATACTGTAGTTTGTTAAATGTGCAATAGCATGCTATCTAA 7148

Qy       1157 AAAAA--CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTT 1214

Db       7149 GAAAAGGTACATACCTTTATGTAAAAGATACTTTTTGCTAAAAATTGCTAACGATCATTT 7208

Qy       1215 GAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGA 1274

Db       7209 GTGCCTTCACTGAGTGGTAAGTGTTTTACTGGTGGAGGGTCTTGCCTGGATGTCGATGGC 7268

Qy       1275 T-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA- 1322

Db       7269 TGCTGACTGATCCGAATGGTGGTTGCCGAAGGCTGAGGTGGCTGTGGCAATTTATTAAAG 7328

Qy       1323 ----ACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTC 1378

Db       7329 TAAGACAACGAGGAAGTTTGTCACATAGATTGACTATTCCTTTCAAG-AAGATTTATATA 7387

Qy       1379 TAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGG-A 1437

Db       7388 TAGCCTGTAATGCTGTTTGGTAGCATTTTGTCTGCGATAGAACTTCTTTCAATATTGGAA 7447

Qy       1438 TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCC 1497

Db       7448 TCAATCCTCTCAAGCCCTGCCACTGCTTTATCGACTAAATTTATATA--ATTCTAAATCC 7505

Qy       1498 ATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTT 1557

Db       7506 TTTGTTGTCATTTACACAGTGCTGTGAACATCTTCACTGGGAGTAGATTCCATCTCAAGA 7565

Qy       1558 CCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATC 1617

Db       7566 ------AACCACTCTTTGCTCATCCATAAGAAGCAACACCTCATTCATTCAAGTTTTAGC 7619

Qy       1618 ATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCA-CTTTTAATTCCAGTTC 1676

Db       7620 ATGAGATT------AAAATGTCATAAATTTAGTTACATCTTCAGGCTCCACTTCTGCTTC 7673

Qy       1677 TCTTGCTGTTTCTACCACATCTGTGGTTC-CTTCCTCCATTGAAGTCTTGAACCTCTCCA 1735

Db       7674 TTTTGCTGTTTCTACCACATCTGCACTTCATTTCCTCCACTGACATCTTGGACACCTCAA 7733

Qy       1736 AGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTG- 1794

Db       7734 ACTCATTCAGGAGGGTTGGAGTCCACATCTTCCAAACTCCTGTTAATGTTGATATTTTGA 7793

Qy       1795 --ACCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAA 1852

Db       7794 CCACCTCCCATGAATTACGAATGTTCTTAATGACATCTAGAATGGTGAATCCTTTCCAGA 7853

Qy       1853 AGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGT 1912

Db       7854 ATGTTTTCAATTTATTTTGCCTAGTTCCA-----ACAAAAAAATCACTATCTATGGCAGC 7908

Qy       1913 TATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTG 1972

Db       7909 TGTTTTCTTATGAGATGTATTTCTTAAATAATACGACTTGAAAGTCAAAATTACTCCCTG 7968

Qy       1973 ATCCATTTTCTGCAAAATAGATGTTGTG 2000

Db       7969 ATCCACAAGCTGAAGAGTGGATGTTGTG 7996


RESULT 29
US-09-949-016-11926/c
; Sequence 11926, Application US/09949016
; Patent No. 6812339
```

```
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11926
;   LENGTH: 95750
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(95750)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11926

  Query Match             21.6%;  Score 433;  DB 3;  Length 95750;
  Best Local Similarity   72.9%;  Pred. No. 1.6e-79;
  Matches  721;  Conservative    0;  Mismatches  225;  Indels   43;  Gaps   11;

Qy       1031 CAAATATGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCA 1090
             ||||      |  |  || |||||| |||||||| ||||| |||| | ||||   ||||||
Db      92430 CAAACTACAATAACATGAATATCATAATAAAGTAAGTCATACAAATGTTTT-ATTTCCCA 92372

Qy       1091 GTGCATATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTAT 1150
             ||||||||||||||  || ||  |||| |||||||||||  | ||||||  | ||   |||
Db      92371 GTGCATATAAAAGTCATGTTTCCACTATACTGTAGTCTACTGAGTGTGAGACAGCAGTAT 92312

Qy       1151 GTCTAAAAAAAC-----ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAA 1205
              ||  | | ||||     ||||||||||||||||||| ||| ||  ||||||||
Db      92311 GTTAACAGAAACAATGTACATACCTTAATTTTAAAATACTTAATAGCTAAAAAA------ 92258

Qy       1206 CAATCATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATT 1265
              ||    |      ||||||||||  |||||||||||||| | || | ||||  || | |
Db      92257 -AAATGCTGTCTTCTTCAGTGAGTCATAATCTTTTTGCTGATAGAGGATCTTGCCTCAGT 92199

Qy       1266 GATGACTGATCGGGGGTCA----------GGTGCTGAAGCTTAGGGTGGCTGTGGCAGT 1314
              | ||      |   | | |             || | |||  | |||| ||||||  || |
Db      92198 GTTGGTGACTGCTGACTGATCTTAGTGGTGATTGTTAAAGGCTGGGGTAGCTGTGACAAT 92139

Qy       1315 TTCTTAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTC 1374
              ||||||| |||||| ||| | |||| | |||  ||||||||||||||||  ||||||||
Db      92138 TTCTTAAGACAACAATGATGTTTGCCACATCGATTGACTCTTCCTTTCACAAAAGATTTC 92079

Qy       1375 TCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATT 1434
              | | |||  | |||  || ||||||||||||||||  |||| || |||||||| | ||||
Db      92078 TATGTAGCATATGAGACTGTTTGATAGCATTTTACCCACAATATAACTTCTTTCACAATT 92019

Qy       1435 GGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGA 1493
              |||  ||||||| ||||| | ||||      ||      | ||||||  || | ||||| | |
Db      92018 GGAGACAATCCTTTCAAATCTTGCTGCTGCTTTCTCAAGTAAGTTCATCTCATATTGTAA 91959

Qy       1494 ATCCATTGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTC 1553
              |||  |||||||||||| |||||| ||||||| | |||||||| |||| |||||||
Db      91958 ATCATTTGTTGTCATTTTAACAATGTTCACAGCGCCTTCACCAGTAGTACATTCCAT--- 91902

Qy       1554 ATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTT 1613
              | |   ||       |||||||||||||||||||||| || ||||||||||  ||||||
Db      91901 -TGTAGGGAACCACTGTCTTTGCTCATCCATAAGAAGCAACTCCTCATCTGTGAAAGTTT 91843

Qy       1614 TATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAG 1673
              ||||| |||||  ||| ||| ||| ||   |||||||     || | ||| | |||| ||
Db      91842 TATCACGAGATCACAGAAATTCAG-CACACCTTCAGG-----CTCCGCTTCTCATTCTAG 91789

Qy       1674 TTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTC 1733
              ||||| |||| |||  |||||||||| ||   ||||||||  ||||||||||  |||
Db      91788 TTCTCGTGCTATTTTCACCACATCTGTAGTGATTTCCTCCA-AGAAGTCTTGAACTCCTC 91730

Qy       1734 CAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTT 1793
              |||||||||||||| |||||||  |||||||||||||||||||||| ||  ||||||
Db      91729 AAAGTCATCCATGAGGATTGGAATAAACTTCTTCCAAATTCCTGTTAATGTTGTTATTTT 91670

Qy       1794 GA--CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAA 1851
              ||  | ||||||||||||  ||||||||||||      |  |||| ||||||||||||
Db      91669 GACTCTTCCCATGAATCAAAAATGTTCTTAATGATGTTTAAAATGATGAATCCTTTCCAG 91610

Qy       1852 AAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAG 1911
              |||||||||||||| || | | | |||||||      || |||   |||  || ||| |||
Db      91609 AAGGTTTTCAATTGACCTTGCCAAGATCCA-----GCAAAGGGATCACAATCTATGGCAG 91555

Qy       1912 TTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTT 1971
              || |||||||   || |||||||  | | |||||| |||||||||  | ||||||| ||
Db      91554 CTACAGCCTTACCAAAAGTATTTCCTAAGTAATAAGACTTGAAAGTTAAGATTACTCATT 91495

Qy       1972 GATCCATTTTCTGCAAAATAGATGTTGTG 2000
              |||||||   ||||| ||| |||| ||||
Db      91494 GATCCATGAGCTGCAGAATGGATGCTGTG 91466


RESULT 30
US-09-949-016-13144/c
; Sequence 13144, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13144
;   LENGTH: 112239
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13144

  Query Match             21.5%;  Score 430.8;  DB 3;  Length 112239;
  Best Local Similarity   74.4%;  Pred. No. 4.6e-79;
  Matches  729;  Conservative    0;  Mismatches  207;  Indels   44;  Gaps   13;

Qy       1047 GGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTT 1106
              ||  | ||||||||||  |||||  |   |    ||| | ||||||| || |||||||||
Db      81868 GGCCACCACAATAAAGCAAGTCATAAGAAATATTTCGTTTCCCAGTACACATAAAAGTTA 81809

Qy       1107 TGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAAAC---- 1162
              || ||| |||| | |||||||  | | ||||||||||   ||||||| |||||||
Db      81808 TGTTTACACTATA-TGTAGTCAATGATGTGTGCAATAACATTATGTCTGAAAAAACAAGG 81750

Qy       1163 -ACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATT 1221
              |||||||||||||  |||||  |||||||  |  ||||||||||  |||||  |||||  ||
Db      81749 TACATACCTTAATTAAAAAATACTTTATTGTT--AAAATGCTAAGGATCATCTGAGCCTT 81692

Qy       1222 CAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGA------- 1274
              ||||||  |||||  |||||||||  |||  ||||||  ||||      ||   ||
Db      81691 TGGTGAGTCATAATCCTTTTGCTGGGGGAGGGTCTTGTCTTGACATTGTTGGATACTGAC 81632

Qy       1275 ----TCGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACA--ACA 1328
                  | ||| |     |||  ||||      ||||  |||||||| ||||||||| |  |||
Db      81631 TGATTAGGGTGGTGATTGCCGAAGGCCGAGGTGACTGTGGCAATTTCTTAAAATAAGACA 81572
```

```
Qy       1329 GTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGA 1388
Db      81571 ATGAAGTTTGTCGCATCAATTGACTCTTCCTTTCGCAACAGATTTCTCTGTAGCATGTGA 81512

Qy       1389 TGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAATCCTCT 1447
Db      81511 TGGTGTTTGATAGCATTTTACCTACAGTAGAACTTCTTTCAAAATTGCAGTCAATCCCGT 81452

Qy       1448 CAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCA 1507
Db      81451 CAAATCCTGCTGCTGCTTTATCAACTAAGTTGATGTAATATTCTAAATTCTTTG-TGTCA 81393

Qy       1508 TTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGG 1567
Db      81392 TCTCAACAATGTTTGCAACATCTTCATTAGGAGTAGATTCCATCTCAAGAAACCACTT-- 81335

Qy       1568 AATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGC 1627
Db      81334 --TCTTTACTCTTCCATAGGAAGCAACTCCTCATTCATTAAAGTTTTATTATGAAGTTGT 81277

Qy       1628 AGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTT 1687
Db      81276 AGCAATTCAGTCACATCTTCAAG-----CTCTATTTCTAATTCTAGTTCTCTTGCTATTT 81222

Qy       1688 CTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGA 1747
Db      81221 CCACCACATCTGCAGTTACTTTCTTCACTAGGATCTTGAACTTCTCAAAGTCATCCATGA 81162

Qy       1748 GGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGACCTC---CCATG 1804
Db      81161 GGGCTGGAATCAACTTCTTCCAAACTCCCATTAATGTTGATATTTTTACTTCTTTCCACG 81102

Qy       1805 AATCATGAATGTTCTTAATGGCA---CCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCA 1861
Db      81101 AAACACAAATGTTCTTAATGGCATCCACCAGAAGGGTGACTCCTTTCTAGAAGGGTTTTG 81042

Qy       1862 ATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTT 1921
Db      81041 ATTTACTTTGCCCAAATCCAT-----CAGAGGAATCACTGTCTGTGACAGCTATAGCTTT 80987

Qy       1922 ATGGAATGTATTTCTTCAATAATAAGGCTTG-AAAGTTGAAATTACTCCTTGATCCATTT 1980
Db      80986 ATGAAATGTATTTCTTAAATAATAGGACTTGAAAAGTTGAAATTACTTCTTGATCCATAG 80927

Qy       1981 TCTGCAAAATAGATGTTGTG 2000
Db      80926 GCTACAGAACGGATGTTGTG 80907
```

RESULT 31
US-09-949-016-13498/c
; Sequence 13498, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13498
; LENGTH: 343352
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(343352)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13498

Query Match 21.4%; Score 429; DB 3; Length 343352;
Best Local Similarity 72.7%; Pred. No. 1.4e-78;
Matches 774; Conservative 0; Mismatches 215; Indels 76; Gaps 14;

```
Qy        977 TACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATA 1036
Db     251976 TACAGGCATACATTGGAGATATTGCAGGTTGGGTTTTCAACCACTGCAATA--------- 251926

Qy       1037 TGCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCAT 1096
Db     251925 ------AAGTGAATACCATAATAAAGTGACTTACACAAATATTTTTGTTTCCCAGTGCAT 251872

Qy       1097 ATAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAA 1156
Db     251871 ACAAAAGCTATGTTTACAGTATACTGTAGTCTATTAAGTGTGCAATAGCACTAGTTCTAA 251812

Qy       1157 AAAAA------CACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATC 1210
Db     251811 ATAAAGCAATGCACATATCTTAATTTAAAAATACTCTATTTCTAAAAAATGCCAAAAATC 251752

Qy       1211 ATTTGAGCATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGA 1270
Db     251751 ATTGGACCCTTCAGTAAGTCATAATCTTTTTGCTGGAAGAAGGTCTTGCCCCCATGTTGA 251692

Qy       1271 CTGAT-----------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCT- 1318
Db     251691 TGGTTGCTGACCGATCAGCATGGTGATTGCTGAAGCATAGGGTGGCTGTGACAGTTTCTT 251632

Qy       1319 ----TAAAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTC 1374
Db     251631 GAAATAACACAACAATGAAGTTGGCCACATTCATTGACTGTTCCTTTCATTCAAGATTTC 251572

Qy       1375 TCTCTAGTGTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATT 1434
Db     251571 TCTGTAACATGTGAGGTTGTTTGATAGCATTTTACCCAGAGTAGAATTTCTTTCAAAATA 251512

Qy       1435 GGA-TCAATCCTCTCAAACCCTGCTCTGCTTTAA------CAACCTAAGTTAATATAATA 1487
Db     251511 AGAGTCAATCCTCTCAAACTCTGCTGCTGTTTAATCAATTAAGCTTAAGCTTATGTAATA 251452

Qy       1488 TTCTGAATCCAT-----------TGTTGTCATTTCAACAATTTTCACAGTGTCTTCACCA 1536
Db     251451 TTCTAAATCTTTTTGTTTGTTTATTTGTTTGTTTCAACAATGCTCACAGGATCTTTACCA 251392

Qy       1537 GGAGTAGATTCCATCTCATTTCCTGAGATGGAATCTTTGCTCATCCATAAGAAGAAATTC 1596
Db     251391 GGAGTAGATGCTATC----TTAAGAAACTACTTTCTTTGATTATCCATAAGAAGCAACTC 251336

Qy       1597 CTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAATACAGTCATGTCTTCAGGCCTCAC 1656
Db     251335 CTCATCTGTTCAAGTTTTATCATAAGATTGCAGCAATTCAGTCACATCTTCAGG-----C 251281

Qy       1657 TTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACCACATCTGTGGTTCCTTCCTCCATT 1716
Db     251280 TCCAGTTCTGATTCTACTTCTCTTGCTATTTCCA-CACGTCTACAATGACTTCCTCTACT 251222

Qy       1717 GAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTTGGAATCGACTTCTTCCAAATTCCT 1776
Db     251221 GAAGGCTTGAA--TCTCAAAGTCATTCATGAGGGTTGGAATCAACTTCTTCCAAACTCCT 251164

Qy       1777 GTTAATATTTATATTTTGA---CCTCCCATGAATCATGAATGTTCTTAATGGCACCTGGA 1833
Db     251163 GTTAATTTTGCTATTTTGATCTTCTCCCATGAATCACAAATGTTTATAACGGTATATGGA 251104

Qy       1834 ATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTTAGTCCAGATCCATCCATCCAGAGG 1893
Db     251103 ATACTGTATTCTTTCCGGAAGGTTTTCAATTTACCTTTTTCTGAT-----CAAGCAGAAC 251049
```

```
Qy       1894 ATCCACTTTCAATGCCAGTTATAGCCTTATGGAATGTATTTCTTCAATAATAAGGCTTGA 1953
Db     251048 AATTAGTAGCTATGGCAGCCACAGCCTTGGGAAATGTATTCCTGAAAT-ATAAGACTTGA 250990

Qy       1954 AAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAATAGATGTTG 1998
Db     250989 AAGTCAAAATTACTCCTTGATCTGTCAGCTGCAAAATGGATGTTG 250945


RESULT 32
US-09-949-016-11852
; Sequence 11852, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11852
;   LENGTH: 321022
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(321022)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11852

  Query Match           21.4%;  Score 428.4;  DB 3;  Length 321022;
  Best Local Similarity   73.3%;  Pred. No. 1.8e-78;
  Matches  762;  Conservative    0;  Mismatches  211;  Indels   67;  Gaps   14;

Qy        978 ACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATAT 1037
Db     137722 ACAGGCATACCTTGGAGATATGGTTGGTTTGGTTCCAGACTACCACAATAAAGCAAATAT 137781

Qy       1038 GCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATA 1097
Db     137782 --------------TTGCAGTAAAGTGAGTCACACAAATGTTTTTGTTTCCCGGTGCATA 137827

Qy       1098 TAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAA 1157
Db     137828 TGAAAGTTATGTCTATACTATACTGTAGTCTGTTAAGTGTCCAATAGCATTATGTCTAAA 137887

Qy       1158 AAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAG 1217
Db     137888 AATGCACATACCTTAATTAAAAATACTTTATTACT-AAAAATGCTAATGATCATCTGAG 137946

Qy       1218 CATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-- 1275
Db     137947 CCTTCAGCAGGTCATAATCTTTTTGGTGGTGAGGGGCTTTGCCTTGATGTTGATGGCTGC 138006

Qy       1276 ------------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA 1323
Db     138007 TGACCGATCACAGTAGTGGTTGCCGAAGAAGGCTGGGGTGGCCGTGGCAATTTCTTAAAA 138066

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTG 1383
Db     138067 CAA---GGAAGTTTGCCACATGA-TAGATTCTTCCTTTCATGAAAGATTTCTCTGTAGCA 138122

Qy       1384 TGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAAT 1442
Db     138123 TGGGATGCTGTTTGACCGCATTTTACCCACAGTAGAACTTCTCTCAAAATTGGAGTCAAT 138182

Qy       1443 CCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGT 1502
Db     138183 CCTCTCAAACCCTGCCACTGCTTTATCAACTATGTTTATGTGATATTCTAAACCCTCTGT 138242

Qy       1503 TGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGA 1562
Db     138243 T-TCATTTCAACA----------GTGTTCACCAGGAGTAGATTCCATATCAGGAAACCA 138290

Qy       1563 GATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAG 1622
Db     138291 CTT----TCTTTGCTCATCCATAAGAAGCAACCCCTCATTTATTCAAGTTTTATCTTGAG 138346

Qy       1623 ATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGC 1682
Db     138347 ATTGCAGCAATTCAGTTCTACCTTCAATCTTCACTTC------TAATACTTATTCTTTGC 138400

Qy       1683 TGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATC 1742
Db     138401 TATTTCCACCATGTCTACAGTTGCTTCCTCTACTGAGGTCTTGAACCCCTCAAAGTCATC 138460

Qy       1743 CATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC---CTC 1799
Db     138461 CATGAGGGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGAAATGTTGACTTTCTT 138520

Qy       1800 CCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTT 1859
Db     138521 TAATGAATCACAAATGTTCTTAATGACATCCAGAATGGTGAATTC-CTCCCAGAGGTTTT 138579

Qy       1860 CAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCC 1919
Db     138580 CAATTTACTTTGCCCAGATCTGT-----AAGAGGAATTGCTATCTATGGGAGCTGGAGAC 138634

Qy       1920 TTATGGAATGTATTTCTTCAATAATAAGGC--TTGAAAGTTGAAATTACTCCTTGATCCA 1977
Db     138635 TTATGAAATATATTTATTAAATCATAAGGCTTTTCAGAGTCAAAATTACTCTTTGATTCA 138694

Qy       1978 TTTTCTGCAAAATAGATGTT 1997
Db     138695 TGGGCTGCAGAATGGATATT 138714


RESULT 33
US-09-949-016-14166
; Sequence 14166, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14166
;   LENGTH: 321022
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(321022)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14166

  Query Match           21.4%;  Score 428.4;  DB 3;  Length 321022;
```

```
  Best Local Similarity   73.3%;  Pred. No. 1.8e-78;
  Matches  762;  Conservative    0;  Mismatches  211;  Indels   67;  Gaps   14;

Qy        978 ACAGGCATACCTTGGAGATACTGTGGGTTTGGTTCCATACCACCACAATAATACAAATAT 1037
              ||||||||||||||||||  || ||||||||||| || |||||||||  |||||||
Db     137722 ACAGGCATACCTTGGAGATATGGTTGGTTTGGTTCCAGACTACCACAATAAAGCAAATAT 137781

Qy       1038 GCAAGAAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATA 1097
                 |  || |||||||||||||||| | |||| | ||||| |||||||
Db     137782 --------------TTGCAGTAAAGTGAGTCACACAAATGTTTTTGTTTCCCGGTGCATA 137827

Qy       1098 TAAAAGTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAA 1157
              | |||||| ||  ||||||| |||||||||||||||||| ||||||  ||||||||||
Db     137828 TGAAAGTTATGTCTATACTATACTGTAGTCTGTTAAGTGTCCAATAGCATTATGTCTAAA 137887

Qy       1158 AAAACACATACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAG 1217
              |||         | | | || ||||| |||||||||| ||||||||||  ||||| ||||
Db     137888 AAATGCACATACCTTAATTAAAAATACTTTATTACT-AAAAATGCTAATGATCATCTGAG 137946

Qy       1218 CATTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT-- 1275
              | |||||   ||  |||||||||||| |||||   ||  ||  |||  || |||  | |
Db     137947 CCTTCAGCAGGTCATAATCTTTTTGGTGGTGAGGGGCTTTGCCTTGATGTTGATGGCTGC 138006

Qy       1276 ------------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAA 1323
                         | |  |  |  ||||  | ||||||| |||||| ||||||||||
Db     138007 TGACCGATCACAGTAGTGGTTGCCGAAGAAGGCTGGGGTGGCCGTGGCAATTTCTTAAAA 138066

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTG 1383
              |||   |||| |||| | || | | || ||||||||||||||||||||||||||| |||
Db     138067 CAA---GGAAGTTTGCCACATGA-TAGATTCTTCCTTTCATGAAAGATTTCTCTGTAGCA 138122

Qy       1384 TGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAAT 1442
              || |||||| |||||  |||||||| |||||||||||||| | |||||||||  |||||
Db     138123 TGGGATGCTGTTTGACCGCATTTTACCCACAGTAGAACTTCTCTCAAAATTGGAGTCAAT 138182

Qy       1443 CCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGT 1502
              ||||||||||||||      || |  | ||| ||| || | ||||||| || ||  |||
Db     138183 CCTCTCAAACCCTGCCACTGCTTTATCAACTATGTTTATGTGATATTCTAAACCCTCTGT 138242

Qy       1503 TGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGA 1562
              | ||||||||||          || ||||||||||||||||||| |||     |
Db     138243 T-TCATTTCAACA----------GTGTTCACCAGGAGTAGATTCCATATCAGGAAACCA 138290

Qy       1563 GATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAG 1622
                |    |||||||||||||||||||||| ||  |||||| | |||||||||||||| ||||
Db     138291 CTT----TCTTTGCTCATCCATAAGAAGCAACCCCTCATTTATTCAAGTTTTATCTTGAG 138346

Qy       1623 ATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGC 1682
              |||||||||||  ||||  |  |||||  | |||||||          | | |  |  ||||
Db     138347 ATTGCAGCAATTCAGTTCTACCTTCAATCTTCACTTC------TAATACTTATTCTTTGC 138400

Qy       1683 TGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATC 1742
              | |||| ||||  |||   ||| |||||||| | ||| |||||||||| ||| |||||||||
Db     138401 TATTTCCACCATGTCTACAGTTGCTTCCTCTACTGAGGTCTTGAACCCCTCAAAGTCATC 138460

Qy       1743 CATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGAC---CTC 1799
              |||||||||||||||| |||||||||||| |||||||||| || | || |||||    ||
Db     138461 CATGAGGGTTGGAATCAACTTCTTCCAAACTCCTGTTAATGTTGAAATGTTGACTTTCTT 138520

Qy       1800 CCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTT 1859
                |||||||  |||||||||||||| || |  |||||||||| |  ||| | |||||||
Db     138521 TAATGAATCACAAATGTTCTTAATGACATCCAGAATGGTGAATTC-CTCCCAGAGGTTTT 138579

Qy       1860 CAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCC 1919
              |||||||||| | |||||||  |       ||||||     || || |||  || |  || |
Db     138580 CAATTTACTTTGCCCAGATCTGT-----AAGAGGAATTGCTATCTATGGGAGCTGGAGAC 138634

Qy       1920 TTATGGAATGTATTTCTTCAATAATAAGGC--TTGAAAGTTGAAATTACTCCTTGATCCA 1977
              ||||| ||| ||||| || ||| |||||||  || | |||  |||||||||| ||||| ||
Db     138635 TTATGAAATATATTTATTAAATCATAAGGCTTTTCAGAGTCAAAATTACTCTTTGATTCA 138694

Qy       1978 TTTTCTGCAAAATAGATGTT 1997
                |  ||||| ||| ||| ||

Db     138695 TGGGCTGCAGAATGGATATT 138714


RESULT 34
US-09-949-016-16419
; Sequence 16419, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16419
;   LENGTH: 61461
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16419

  Query Match            21.4%;  Score 427.8;  DB 3;  Length 61461;
  Best Local Similarity   73.8%;  Pred. No. 1.7e-78;
  Matches  717;  Conservative    0;  Mismatches  212;  Indels   42;  Gaps   12;

Qy       1051 ATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAAGTTTTGCT 1110
              | ||||||||||  ||||||| || | ||||| ||||||||||||||| |||||| || |
Db     39855 ACCACAATAAAGCAAGTCACATAATTTTTTTGTCTTCCCAGTGCATATGAAAGTTATGTT 39914

Qy       1111 TATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA----ACACAT 1166
              | | ||| |||||||||| ||    ||||||| ||  ||||| ||||||||       | ||
Db     39915 TTTGCTATACTGTAGTCTATT--TTGTGCAACAGCATTATGCCTAAAAAACAATGTATAT 39972

Qy       1167 ACCTTAATTTTAAAATGCTTTATTACTAAAAAATGCTAACAATCATTTGAGCATTCAGTG 1226
              |||||||||  |||||  | |||  | ||||||||||||||  ||||  ||||  ||||| |
Db     39973 ACCTTAATTAAAAAATACGTTACTCCTAAAAAATGCTAATGATCACCTGAGTCTTCAGCG 40032

Qy       1227 AGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTC----------TTATTGATGACTGATC 1276
              ||||||| |||||||||||||||| |||||||  |          | ||| |||||||||
Db     40033 AGTTGTAGTCTTTTTGCTGGTGGAGGGTCTTTCCTAGGTGTTGATGGTTGCTGACTGATC 40092

Qy       1277 GGGGGTCAGGT--GCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAAACA--ACAGTGA 1332
              |||   |||     | | ||| |||||||||| |  |||||||| |  ||| |||
Db     40093 AGGGTAGTGGTTGCTGAAGGTTTAAGGTGGCTGTGGTAACTTCTTAAAATAAGACAATGA 40152

Qy       1333 AGATTGCAATATCAGTTGACTCTTCCTTTCATGAAAGATTTCTCTCTAGTGTGTGATGCT 1392
              || |||    |||  || ||| |||||||| |||||||||||||| ||   || || |||
Db     40153 AGTTTGATGCATCGATTAACTTTTCCTTTCAGGAAAGATTTCTCTGTACCATGAGAAGCT 40212

Qy       1393 TTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGATCAATCCTCTCAAAC 1452
              |||||||||||||||  |||||||  ||||| ||  | | | |  ||| ||||||||||
Db     40213 ATTTGATAGCATTTTACCCACAGTACAACTTTTTCAACATTGGAGTCAGTCCTCTCAAA- 40271

Qy       1453 CCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTGTTGTCATTTCA 1512
              ||||       || |       |||||  | |||||||| |||||  |||||||  ||||||
Db     40272 -CTGCCACTGCTTTATCTGTGAAGTTTGTGTAATATTCTAAATCCTTTGTTGTTATTTCA 40330

Qy       1513 ACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTGAGATGGAATCT 1572
              |||| |||||||  |||||||||||||| |||||||| ||        | |    |||
Db     40331 GCAATGTTCACAGCATCTTCACCAGGAGTTGATTCCATCCCAAGAAACCACTT----TCT 40386

Qy       1573 TTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGAGATTGCAGCAA 1632
              |||||||| |||||||||| || ||||    || ||||||||||     ||||||||||
Db     40387 TTGCTCAT-CATAAGAAGAAACTCTTCATTCATTTAAGTTTTATCTCATGATTGCAGCAA 40445

Qy       1633 TACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTGCTGTTTCTACC 1692
```

```
                |  |||||   ||||||||  |||  ||       | | | ||| |||| ||||||||  ||
Db      40446 TTTAGTCACAGCTTCAGGCTCCACATC------TAATTCGGTTATCTTACTGTTTCTGCC 40499

Qy       1693 ACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCATCCATGAGGGTT 1752
              ||| | |  ||  ||| |||||    ||||||||||  ||| ||||||||||||||||||
Db      40500 ACAGCAGCAGTGACTTTCTCCACCATAGTCTTGAACTCCTCAAAGTCATCCATGAGGGTT 40559

Qy       1753 GGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCTCCCATGAATCA 1809
              |||||| || ||||||||||||||||||||  | |||||||||    |||||  | |||||
Db      40560 GGAATCAACATCTTCCAAATTCCTGTTAATGCTGATATTTTGACCTCCTCCTGTAAATCA 40619

Qy       1810 TGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTTTCAATTTACTT 1869
                |||||||||||||||| || | |||||||||||||||||| || ||||||||||| ||
Db      40620 CAAATGTTCTTAATGGCATCTAG-ATGGTGAATCCTTTCCAGAAAGTTTTCAATTTATTT 40678

Qy       1870 AGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGCCTTATGGAATG 1929
              |  |  || |||     ||||| |  || |  | ||  ||| ||||   ||||  |||
Db      40679 TGCTCTTATTCAT-----CAGAGAAATCATTAACTATAGCAGCTATAATGTTATTAAATA 40733

Qy       1930 TATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCATTTTCTGCAAAA 1989
              |||||||| | ||||||| |||||||||  |||||||||||||||||||  ||    |
Db      40734 TATTTCTTAACTAATAAGACTTGAAAGTCAAAATTACTCCTTGATCCATGGGCTATGGAG 40793

Qy       1990 TAGATGTTGTG 2000
              | || |||||||
Db      40794 TGGAGGTTGTG 40804


RESULT 35
US-09-949-016-14546/c
; Sequence 14546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14546
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14546

  Query Match             21.3%;  Score 426;  DB 3;  Length 818128;
  Best Local Similarity   74.0%;  Pred. No. 6.8e-78;
  Matches  727;  Conservative    0;  Mismatches  210;  Indels   45;  Gaps   13;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
              ||||  |||||||||||||| |||| ||||| | | |||| | |||||         ||||
Db     193812 AAGTAAATATCACAATAAAATGAGGCACACGAATATTTTTGGTTCCCCAGTGCATTAAAT 193753

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA-A 1161
              ||| || ||| |||  ||| |  ||||| ||||||||||||||  || | |||||||||
Db     193752 GTTATGTTTACACTCTACTATCATCTGTCAAGTGTGCAATAGCATTGTATCTAAAAAATG 193693

Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACT--AAAAAATGCTAACAATCATTTGAGCA 1219
              |||||||||||||| ||||| | ||| | ||  |||||||| |||||||||| ||||
Db     193692 TACATACCTTAATTTGAAAATACATTACTCCTAAAAAAAATGGTAACAATCATCTGAGCC 193633

Qy       1220 TTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---- 1275
              | ||| |||| ||||||||||||||||  | | ||||  || | ||| ||  | |
Db     193632 TGCAGGGAGTCATAATCTTTTTGCTGGTACAGGCTCTTGCCTCAGTGAGGATGGCTGCTG 193573

Qy       1276 -------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----A 1323
                    ||| | | | |||||||| |  || ||||||| ||| |||||||||      |
Db     193572 ATTGATCAGGGTGGCGGTTGCTGAAGGTCGGGATGGCTGTAGCAATTTCTTAAAATAAGA 193513

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT-TCATGAAAGATTTCTCTCTAGT 1382
              ||||| |||| ||||   |||  ||| |||| | || |    || ||||||| | |||
Db     193512 CAACAACGAAGTTTGCTGCATCGATTGGCTCTCCTTTATAGCCAATGATTTCTTTGTAGC 193453

Qy       1383 GTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAA 1441
              |||||||| |||||||||||||||  | |||||||||||||| |||||| || ||||
Db     193452 ATGTGATGCTGTTTGATAGCATTTTA--CTCAGTAGAACTTCTTTCAAAATTTGAGTCAA 193395

Qy       1442 TCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTG 1501
              || |||||||||||| |     || |  | ||||||| ||  |||||||| ||||| |||
Db     193394 TCTTCTCAAACCCTGATGCTAATTTATCAACTAAGTTTATGGAATATTCTAAATCCTTTG 193335

Qy       1502 TTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTG 1561
              |||||||||||||||| |||||||  || ||||||  |||||||  | ||||||||
Db     193334 TTGTCATTTCAACAATGTTCACAGCATCATCACCAAGAGTAGAGTGCATCTCATGAAACC 193275

Qy       1562 AGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGA 1621
              |     |||||||||||||||||| || || ||   ||| | | |||||||| | |||
Db     193274 AC----AATCTTTGCTCATCCATAAGCAGCAACTCTGGATCCGCTGAAGTTTTACCTTGA 193219

Qy       1622 GATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTG 1681
              |||| |||| || ||||||   |||    || ||| ||||| | ||||  ||||||  ||
Db     193218 GATTACAGCTATTCAGTCACACCTT-----CTGACTCCACTTCTCATTCTGGTTCTCATG 193164

Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741
              || |||||||||||||||  ||  |||| || | |||||||||| ||| ||| |||||||
Db     193163 CTATTTCTACCACATCTGCAGTGACTTCTTCTACTGAAGTCTTGGACC-CTCAAAGTCAT 193105

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCT 1798
              |||||  ||||||||  ||| |||||||| |||||  |||| || |||||||||    |||
Db     193104 TGATGAGTGTTGGAATTAACTACTTCCAAACTCCTGATAATGTTGATATTTTGACCTCCT 193045

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858
              ||| | |||||||||| ||||| |||||  ||||||||||||   ||||||     ||||
Db     193044 CCCCTCAATCATGAATTTTCTTCATGGCATCTGGAATGGTGAGGACTTTCC-----GTTT 192990

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
              ||| |||||||  |||||||||||| |      |  ||  |     ||  ||| |||| |
Db     192989 TCAGTTTACTTTATCCAGATCCATCAGT----GGACTCACCACAATATAGCAGCTATACC 192934

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              |||||  |||||||   || | ||||||| ||||||||||||| | ||| |||||||||
Db     192933 TTTATGAAATGTATGATTTAAGTAATAAGACTTGAAAGTTGAAGTGACTTCTTGATCCAA 192874

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000
                 ||||| ||| |||||||||
Db     192873 GGACTGCAGAATGGATGTTGTG 192852


RESULT 36
US-09-949-016-14547/c
; Sequence 14547, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
```

```
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14547
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14547

  Query Match             21.3%;  Score 426;  DB 3;  Length 818128;
  Best Local Similarity   74.0%;  Pred. No. 6.8e-78;
  Matches  727;  Conservative    0;  Mismatches  210;  Indels   45;  Gaps   13;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
             ||||  |||||||||||||| |||| ||||| | | |||| | |||||        ||||
Db     193812 AAGTAAATATCACAATAAAATGAGGCACACGAATATTTTTGGTTCCCCAGTGCATTAAAT 193753

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA-A 1161
             ||| || ||| |||  ||| |  ||||| |||||||||||||  || | |||||||||
Db     193752 GTTATGTTTACACTCTACTATCATCTGTCAAGTGTGCAATAGCATTGTATCTAAAAAATG 193693

Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACT--AAAAAATGCTAACAATCATTTGAGCA 1219
              |||||||||||||| ||||| | ||| | ||  |||||||| |||||||||| |||||
Db     193692 TACATACCTTAATTTGAAAATACATTACTCCTAAAAAAAATGGTAACAATCATCTGAGCC 193633

Qy       1220 TTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---- 1275
             | ||| ||||  ||||||||||||||||  | | ||||  || | ||| ||  | |
Db     193632 TGCAGGGAGTCATAATCTTTTTGCTGGTACAGGCTCTTGCCTCAGTGAGGATGGCTGCTG 193573

Qy       1276 -------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----A 1323
                   ||| | | | |||||||| |  || ||||||| ||| |||||||||     |
Db     193572 ATTGATCAGGGTGGCGGTTGCTGAAGGTCGGGATGGCTGTAGCAATTTCTTAAAATAAGA 193513

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT-TCATGAAAGATTTCTCTCTAGT 1382
             |||||  |||| ||||   |||  ||| |||| | || |    || |||||||| | |||
Db     193512 CAACAACGAAGTTTGCTGCATCGATTGGCTCTCCTTTATAGCCAATGATTTCTTTGTAGC 193453

Qy       1383 GTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAA 1441
              |||||||| ||||||||||||||  | ||||||||||||||| |||||| || ||||
Db     193452 ATGTGATGCTGTTTGATAGCATTTTA--CTCAGTAGAACTTCTTTCAAAATTTGAGTCAA 193395

Qy       1442 TCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTG 1501
             || |||||||||||| |      || |  | |||||||| ||  |||||||| ||||| |||
Db     193394 TCTTCTCAAACCCTGATGCTAATTTATCAACTAAGTTTATGGAATATTCTAAATCCTTTG 193335

Qy       1502 TTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTG 1561
              |||||||||||||||| |||||||  || ||||||  |||||||  | ||||||||
Db     193334 TTGTCATTTCAACAATGTTCACAGCATCATCACCAAGAGTAGAGTGCATCTCATGAAACC 193275

Qy       1562 AGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGA 1621
             |    ||||||||||||||||||||| || || ||   ||| | | ||||||||  | |||
Db     193274 AC----AATCTTTGCTCATCCATAAGCAGCAACTCTGGATCCGCTGAAGTTTTACCTTGA 193219

Qy       1622 GATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTG 1681
              |||| |||| || ||||||   |||     || ||| ||||| | ||||  |||||| ||
Db     193218 GATTACAGCTATTCAGTCACACCTT-----CTGACTCCACTTCTCATTCTGGTTCTCATG 193164

Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741
              || ||||||||||||||   ||  |||| || | |||||||||| ||| ||| |||||||
Db     193163 CTATTTCTACCACATCTGCAGTGACTTCTTCTACTGAAGTCTTGGACC-CTCAAAGTCAT 193105

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCT 1798
               ||||| ||||||||  ||| ||||||||| ||||| |||| || |||||||||   |||
Db     193104 TGATGAGTGTTGGAATTAACTACTTCCAAACTCCTGATAATGTTGATATTTTGACCTCCT 193045

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858
              ||| | |||||||||| ||||| |||||| |||||||||||    ||||||     ||||
Db     193044 CCCCTCAATCATGAATTTTCTTCATGGCATCTGGAATGGTGAGGACTTTCC-----GTTT 192990

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
              ||| |||||||  |||||||||||  |     |  ||  |      ||  ||| |||| |
Db     192989 TCAGTTTACTTTATCCAGATCCATCAGT----GGACTCACCACAATATAGCAGCTATACC 192934

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              ||||| |||||||   || | ||||||| |||||||||||| | ||| |||||||||
Db     192933 TTTATGAAATGTATGATTTAAGTAATAAGACTTGAAAGTTGAAGTGACTTCTTGATCCAA 192874

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000
                 ||||| ||| |||||||||
Db     192873 GGACTGCAGAATGGATGTTGTG 192852


RESULT 37
US-09-949-016-14548/c
; Sequence 14548, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14548
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14548

  Query Match             21.3%;  Score 426;  DB 3;  Length 818128;
  Best Local Similarity   74.0%;  Pred. No. 6.8e-78;
  Matches  727;  Conservative    0;  Mismatches  210;  Indels   45;  Gaps   13;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
             ||||  |||||||||||||| |||| ||||| | | |||| | |||||        ||||
Db     193812 AAGTAAATATCACAATAAAATGAGGCACACGAATATTTTTGGTTCCCCAGTGCATTAAAT 193753

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA-A 1161
             ||| || ||| |||  ||| |  ||||| |||||||||||||  || | |||||||||
Db     193752 GTTATGTTTACACTCTACTATCATCTGTCAAGTGTGCAATAGCATTGTATCTAAAAAATG 193693

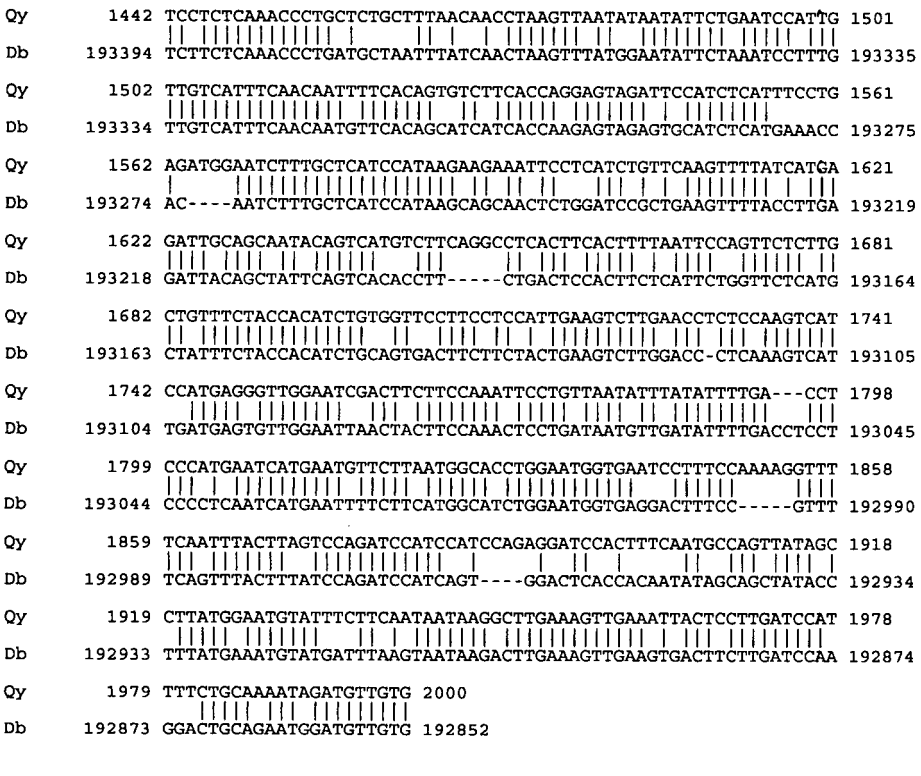
Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACT--AAAAAATGCTAACAATCATTTGAGCA 1219
              |||||||||||||| ||||| | ||| | ||  |||||||| |||||||||| |||||
Db     193692 TACATACCTTAATTTGAAAATACATTACTCCTAAAAAAAATGGTAACAATCATCTGAGCC 193633

Qy       1220 TTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---- 1275
             | ||| ||||  ||||||||||||||||  | | ||||  || | ||| ||  | |
Db     193632 TGCAGGGAGTCATAATCTTTTTGCTGGTACAGGCTCTTGCCTCAGTGAGGATGGCTGCTG 193573

Qy       1276 -------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----A 1323
                   ||| | | | |||||||| |  || ||||||| ||| |||||||||     |
Db     193572 ATTGATCAGGGTGGCGGTTGCTGAAGGTCGGGATGGCTGTAGCAATTTCTTAAAATAAGA 193513

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT-TCATGAAAGATTTCTCTCTAGT 1382
             |||||  |||| ||||   |||  ||| |||| | || |    || |||||||| | |||
Db     193512 CAACAACGAAGTTTGCTGCATCGATTGGCTCTCCTTTATAGCCAATGATTTCTTTGTAGC 193453

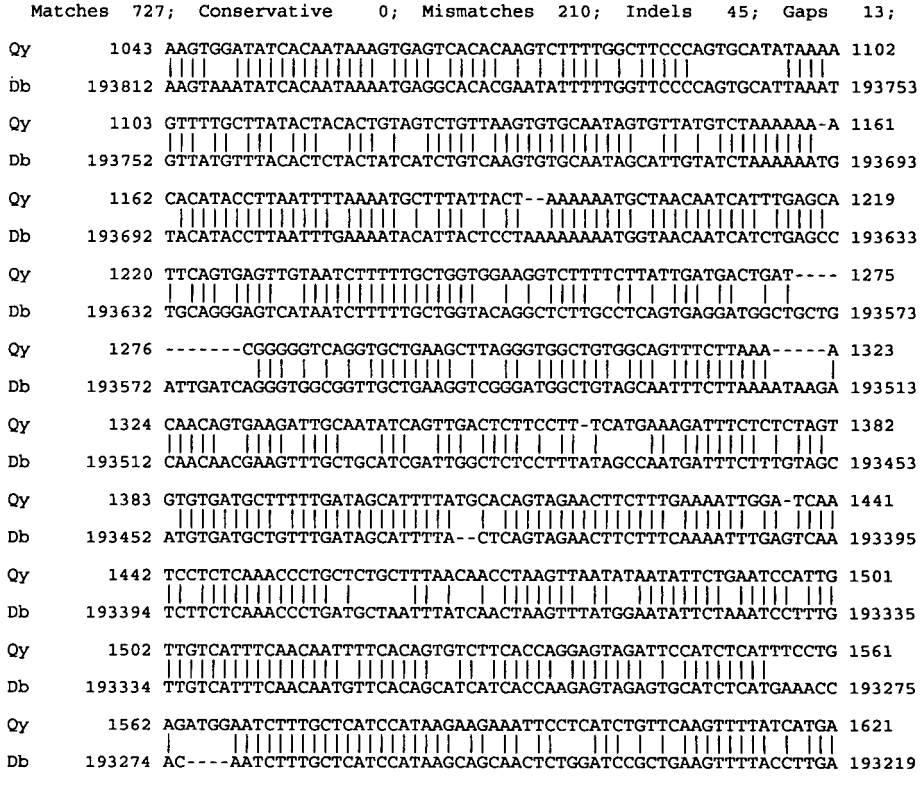
Qy       1383 GTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAA 1441
              |||||||| ||||||||||||||  | ||||||||||||||| |||||| || ||||
Db     193452 ATGTGATGCTGTTTGATAGCATTTTA--CTCAGTAGAACTTCTTTCAAAATTTGAGTCAA 193395
```



RESULT 38
US-09-949-016-14549/c
; Sequence 14549, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14549
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14549

Query Match 21.3%; Score 426; DB 3; Length 818128;
Best Local Similarity 74.0%; Pred. No. 6.8e-78;

Matches 727; Conservative 0; Mismatches 210; Indels 45; Gaps 13;



RESULT 39
US-09-949-016-14550/c

```
; Sequence 14550, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14550
;  LENGTH: 818128
;  TYPE: DNA
;  ORGANISM: Human
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(818128)
;  OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14550

  Query Match             21.3%;  Score 426;  DB 3;  Length 818128;
  Best Local Similarity   74.0%;  Pred. No. 6.8e-78;
  Matches  727;  Conservative    0;  Mismatches  210;  Indels   45;  Gaps   13;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
             ||||  |||||||||||||| |||| |||||  | | |||| | |||||         ||||
Db     193812 AAGTAAATATCACAATAAAATGAGGCACACGAATATTTTTGGTTCCCCAGTGCATTAAAT 193753

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA-A 1161
             ||| || ||| |||  ||| |  ||||| |||||||||||||  || | |||||||||
Db     193752 GTTATGTTTACACTCTACTATCATCTGTCAAGTGTGCAATAGCATTGTATCTAAAAAATG 193693

Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACT--AAAAAATGCTAACAATCATTTGAGCA 1219
              |||||||||||||| ||||| | ||| | ||  ||||||||  ||||||||| |||||
Db     193692 TACATACCTTAATTTGAAAATACATTACTCCTAAAAAAAATGGTAACAATCATCTGAGCC 193633

Qy       1220 TTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---- 1275
             | ||| ||||  ||||||||||||||||  | | ||||  || | ||| ||  | |
Db     193632 TGCAGGGAGTCATAATCTTTTTGCTGGTACAGGCTCTTGCCTCAGTGAGGATGGCTGCTG 193573

Qy       1276 -------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----A 1323
                    ||| | | | |||||||| |  || ||||||| ||| |||||||||     |
Db     193572 ATTGATCAGGGTGGCGGTTGCTGAAGGTCGGGATGGCTGTAGCAATTTCTTAAAATAAGA 193513

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT-TCATGAAAGATTTCTCTCTAGT 1382
             |||||  |||| ||||   |||  ||| |||| | || |    || |||||||| | |||
Db     193512 CAACAACGAAGTTTGCTGCATCGATTGGCTCTCCTTTATAGCCAATGATTTCTTTGTAGC 193453

Qy       1383 GTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAA 1441
              |||||||| ||||||||||||||  | |||||||||||||| |||||| || ||||
Db     193452 ATGTGATGCTGTTTGATAGCATTTTA--CTCAGTAGAACTTCTTTCAAAATTTGAGTCAA 193395

Qy       1442 TCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTG 1501
             || |||||||||||| |      || |  | ||||||| ||  |||||||| ||||| |||
Db     193394 TCTTCTCAAACCCTGATGCTAATTTATCAACTAAGTTTATGGAATATTCTAAATCCTTTG 193335

Qy       1502 TTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTG 1561
              |||||||||||||||| |||||||  || ||||||  |||||||  | |||||||||
Db     193334 TTGTCATTTCAACAATGTTCACAGCATCATCACCAAGAGTAGAGTGCATCTCATGAAACC 193275

Qy       1562 AGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGA 1621
              |    ||||||||||||||||||||| || ||  ||| | | ||||||||  | |||
Db     193274 AC----AATCTTTGCTCATCCATAAGCAGCAACTCTGGATCCGCTGAAGTTTTACCTTGA 193219

Qy       1622 GATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTG 1681
              |||| |||| || ||||||   |||     || ||| ||||| | ||||  |||||| ||

Db     193218 GATTACAGCTATTCAGTCACACCTT-----CTGACTCCACTTCTCATTCTGGTTCTCATG 193164

Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741
              || |||||||||||||||  ||  |||| || | |||||||||| ||| ||| |||||||
Db     193163 CTATTTCTACCACATCTGCAGTGACTTCTTCTACTGAAGTCTTGGACC-CTCAAAGTCAT 193105

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCT 1798
              |||||  ||||||||  ||| |||||||| |||||  |||| || |||||||||   |||
Db     193104 TGATGAGTGTTGGAATTAACTACTTCCAAACTCCTGATAATGTTGATATTTTGACCTCCT 193045

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858
              ||| | |||||||||| ||||| |||||| ||||||||||||   ||||||     ||||
Db     193044 CCCCTCAATCATGAATTTTCTTCATGGCATCTGGAATGGTGAGGACTTTCC-----GTTT 192990

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
              ||| |||||||  |||||||||||||  |     |  ||  |     ||  ||| |||| |
Db     192989 TCAGTTTACTTTATCCAGATCCATCAGT----GGACTCACCACAATATAGCAGCTATACC 192934

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              |||||  |||||||   || | ||||||| |||||||||||||  | ||| |||||||||
Db     192933 TTTATGAAATGTATGATTTAAGTAATAAGACTTGAAAGTTGAAGTGACTTCTTGATCCAA 192874

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000
                 ||||| ||| |||||||||
Db     192873 GGACTGCAGAATGGATGTTGTG 192852


RESULT 40
US-09-949-016-14551/c
; Sequence 14551, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14551
;  LENGTH: 818128
;  TYPE: DNA
;  ORGANISM: Human
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(818128)
;  OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14551

  Query Match             21.3%;  Score 426;  DB 3;  Length 818128;
  Best Local Similarity   74.0%;  Pred. No. 6.8e-78;
  Matches  727;  Conservative    0;  Mismatches  210;  Indels   45;  Gaps   13;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
             ||||  |||||||||||||| |||| |||||  | | |||| | |||||         ||||
Db     193812 AAGTAAATATCACAATAAAATGAGGCACACGAATATTTTTGGTTCCCCAGTGCATTAAAT 193753

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA-A 1161
             ||| || ||| |||  ||| |  ||||| |||||||||||||  || | |||||||||
Db     193752 GTTATGTTTACACTCTACTATCATCTGTCAAGTGTGCAATAGCATTGTATCTAAAAAATG 193693

Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACT--AAAAAATGCTAACAATCATTTGAGCA 1219
              |||||||||||||| ||||| | ||| | ||  ||||||||  ||||||||| |||||
Db     193692 TACATACCTTAATTTGAAAATACATTACTCCTAAAAAAAATGGTAACAATCATCTGAGCC 193633

Qy       1220 TTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---- 1275
```

```
Db     193632 TGCAGGGAGTCATAATCTTTTTGCTGGTACAGGCTCTTGCCTCAGTGAGGATGGCTGCTG 193573

Qy       1276 -------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----A 1323

Db     193572 ATTGATCAGGGTGGCGGTTGCTGAAGGTCGGGATGGCTGTAGCAATTTCTTAAAATAAGA 193513

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT-TCATGAAAGATTTCTCTCTAGT 1382

Db     193512 CAACAACGAAGTTTGCTGCATCGATTGGCTCTCCTTTATAGCCAATGATTTCTTTGTAGC 193453

Qy       1383 GTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAA 1441

Db     193452 ATGTGATGCTGTTTGATAGCATTTTA--CTCAGTAGAACTTCTTTCAAAATTTGAGTCAA 193395

Qy       1442 TCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTG 1501

Db     193394 TCTTCTCAAACCCTGATGCTAATTTATCAACTAAGTTTATGGAATATTCTAAATCCTTTG 193335

Qy       1502 TTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTG 1561

Db     193334 TTGTCATTTCAACAATGTTCACAGCATCATCACCAAGAGTAGAGTGCATCTCATGAAACC 193275

Qy       1562 AGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGA 1621

Db     193274 AC----AATCTTTGCTCATCCATAAGCAGCAACTCTGGATCCGCTGAAGTTTTACCTTGA 193219

Qy       1622 GATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTG 1681

Db     193218 GATTACAGCTATTCAGTCACACCTT-----CTGACTCCACTTCTCATTCTGGTTCTCATG 193164

Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741

Db     193163 CTATTTCTACCACATCTGCAGTGACTTCTTCTACTGAAGTCTTGGACC-CTCAAAGTCAT 193105

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCT 1798

Db     193104 TGATGAGTGTTGGAATTAACTACTTCCAAACTCCTGATAATGTTGATATTTTGACCTCCT 193045

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858

Db     193044 CCCCTCAATCATGAATTTTCTTCATGGCATCTGGAATGGTGAGGACTTTCC-----GTTT 192990

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918

Db     192989 TCAGTTTACTTTATCCAGATCCATCAGT----GGACTCACCACAATATAGCAGCTATACC 192934

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978

Db     192933 TTTATGAAATGTATGATTTAAGTAATAAGACTTGAAAGTTGAAGTGACTTCTTGATCCAA 192874

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000

Db     192873 GGACTGCAGAATGGATGTTGTG 192852


RESULT 41
US-09-949-016-14552/c
; Sequence 14552, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14552
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14552

  Query Match             21.3%;  Score 426;  DB 3;  Length 818128;
  Best Local Similarity   74.0%;  Pred. No. 6.8e-78;
  Matches  727;  Conservative    0;  Mismatches  210;  Indels   45;  Gaps   13;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102

Db     193812 AAGTAAATATCACAATAAAATGAGGCACACGAATATTTTTGGTTCCCCAGTGCATTAAAT 193753

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA-A 1161

Db     193752 GTTATGTTTACACTCTACTATCATCTGTCAAGTGTGCAATAGCATTGTATCTAAAAAATG 193693

Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACT--AAAAAATGCTAACAATCATTTGAGCA 1219

Db     193692 TACATACCTTAATTTGAAAATACATTACTCCTAAAAAAAATGGTAACAATCATCTGAGCC 193633

Qy       1220 TTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---- 1275

Db     193632 TGCAGGGAGTCATAATCTTTTTGCTGGTACAGGCTCTTGCCTCAGTGAGGATGGCTGCTG 193573

Qy       1276 -------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----A 1323

Db     193572 ATTGATCAGGGTGGCGGTTGCTGAAGGTCGGGATGGCTGTAGCAATTTCTTAAAATAAGA 193513

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT-TCATGAAAGATTTCTCTCTAGT 1382

Db     193512 CAACAACGAAGTTTGCTGCATCGATTGGCTCTCCTTTATAGCCAATGATTTCTTTGTAGC 193453

Qy       1383 GTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAA 1441

Db     193452 ATGTGATGCTGTTTGATAGCATTTTA--CTCAGTAGAACTTCTTTCAAAATTTGAGTCAA 193395

Qy       1442 TCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTG 1501

Db     193394 TCTTCTCAAACCCTGATGCTAATTTATCAACTAAGTTTATGGAATATTCTAAATCCTTTG 193335

Qy       1502 TTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTG 1561

Db     193334 TTGTCATTTCAACAATGTTCACAGCATCATCACCAAGAGTAGAGTGCATCTCATGAAACC 193275

Qy       1562 AGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGA 1621

Db     193274 AC----AATCTTTGCTCATCCATAAGCAGCAACTCTGGATCCGCTGAAGTTTTACCTTGA 193219

Qy       1622 GATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTG 1681

Db     193218 GATTACAGCTATTCAGTCACACCTT-----CTGACTCCACTTCTCATTCTGGTTCTCATG 193164

Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741

Db     193163 CTATTTCTACCACATCTGCAGTGACTTCTTCTACTGAAGTCTTGGACC-CTCAAAGTCAT 193105

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCT 1798

Db     193104 TGATGAGTGTTGGAATTAACTACTTCCAAACTCCTGATAATGTTGATATTTTGACCTCCT 193045

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858

Db     193044 CCCCTCAATCATGAATTTTCTTCATGGCATCTGGAATGGTGAGGACTTTCC-----GTTT 192990

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
```

```
              ||| |||||||  ||||||||||||  |     |  ||  |     ||  ||| |||| |
Db     192989 TCAGTTTACTTTATCCAGATCCATCAGT----GGACTCACCACAATATAGCAGCTATACC 192934

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              |||||  |||||||   ||  | ||||||| |||||||||||||  | |||  |||||||||
Db     192933 TTTATGAAATGTATGATTTAAGTAATAAGACTTGAAAGTTGAAGTGACTTCTTGATCCAA 192874

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000
                |||||  |||  |||||||||
Db     192873 GGACTGCAGAATGGATGTTGTG 192852


RESULT 42
US-09-949-016-14553/c
; Sequence 14553, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14553
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14553

  Query Match             21.3%;  Score 426;  DB 3;  Length 818128;
  Best Local Similarity   74.0%;  Pred. No. 6.8e-78;
  Matches  727;  Conservative    0;  Mismatches  210;  Indels   45;  Gaps   13;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
              ||||  |||||||||||||| |||| ||||| | | |||| | |||||         ||||
Db     193812 AAGTAAATATCACAATAAAATGAGGCACACGAATATTTTTGGTTCCCCAGTGCATTAAAT 193753

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA-A 1161
              ||| || ||| |||  ||| |  ||||| ||||||||||||||  || | |||||||||
Db     193752 GTTATGTTTACACTCTACTATCATCTGTCAAGTGTGCAATAGCATTGTATCTAAAAAATG 193693

Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACT--AAAAAATGCTAACAATCATTTGAGCA 1219
              |||||||||||||| ||||| | ||| | ||  |||||||| |||||||||| |||||
Db     193692 TACATACCTTAATTTGAAAATACATTACTCCTAAAAAAAATGGTAACAATCATCTGAGCC 193633

Qy       1220 TTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---- 1275
              | ||| ||||  ||||||||||||||||  | | ||||  || | ||| ||  | |
Db     193632 TGCAGGGAGTCATAATCTTTTTGCTGGTACAGGCTCTTGCCTCAGTGAGGATGGCTGCTG 193573

Qy       1276 -------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----A 1323
                    ||| | | | |||||||| |  || ||||||| ||| |||||||||     |
Db     193572 ATTGATCAGGGTGGCGGTTGCTGAAGGTCGGGATGGCTGTAGCAATTTCTTAAAATAAGA 193513

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT-TCATGAAAGATTTCTCTCTAGT 1382
              |||||  |||| ||||   |||  ||| |||| | || |    || ||||||| | |||
Db     193512 CAACAACGAAGTTTGCTGCATCGATTGGCTCTCCTTTATAGCCAATGATTTCTTTGTAGC 193453

Qy       1383 GTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAA 1441
              |||||||| ||||||||||||||  | ||||||||||||||| ||||||  || ||||
Db     193452 ATGTGATGCTGTTTGATAGCATTTTA--CTCAGTAGAACTTCTTTCAAAATTTGAGTCAA 193395

Qy       1442 TCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTG 1501
              || |||||||||||| |     || |  | ||||||| ||  |||||||| |||||  |||
Db     193394 TCTTCTCAAACCCTGATGCTAATTTATCAACTAAGTTTATGGAATATTCTAAATCCTTTG 193335

Qy       1502 TTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTG 1561
              ||||||||||||||||| |||||||  || ||||||  |||||||  | ||||||||
Db     193334 TTGTCATTTCAACAATGTTCACAGCATCATCACCAAGAGTAGAGTGCATCTCATGAAACC 193275

Qy       1562 AGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGA 1621
              |     |||||||||||||||||||| || ||  ||   ||| | | |||||||| | |||
Db     193274 AC----AATCTTTGCTCATCCATAAGCAGCAACTCTGGATCCGCTGAAGTTTTACCTTGA 193219

Qy       1622 GATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTG 1681
              |||| |||| || ||||||   |||     || ||| ||||| | ||||  ||||||  ||
Db     193218 GATTACAGCTATTCAGTCACACCTT-----CTGACTCCACTTCTCATTCTGGTTCTCATG 193164

Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741
              || ||||||||||||||  ||  |||| || | |||||||||| ||| ||| |||||||
Db     193163 CTATTTCTACCACATCTGCAGTGACTTCTTCTACTGAAGTCTTGGACC-CTCAAAGTCAT 193105

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCT 1798
              |||||  ||||||||  ||| |||||||| |||||  |||| || |||||||||   |||
Db     193104 TGATGAGTGTTGGAATTAACTACTTCCAAACTCCTGATAATGTTGATATTTTGACCTCCT 193045

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858
              ||| | |||||||||| |||||  ||||||  ||||||||||||   ||||||     ||||
Db     193044 CCCCTCAATCATGAATTTTCTTCATGGCATCTGGAATGGTGAGGACTTTCC-----GTTT 192990

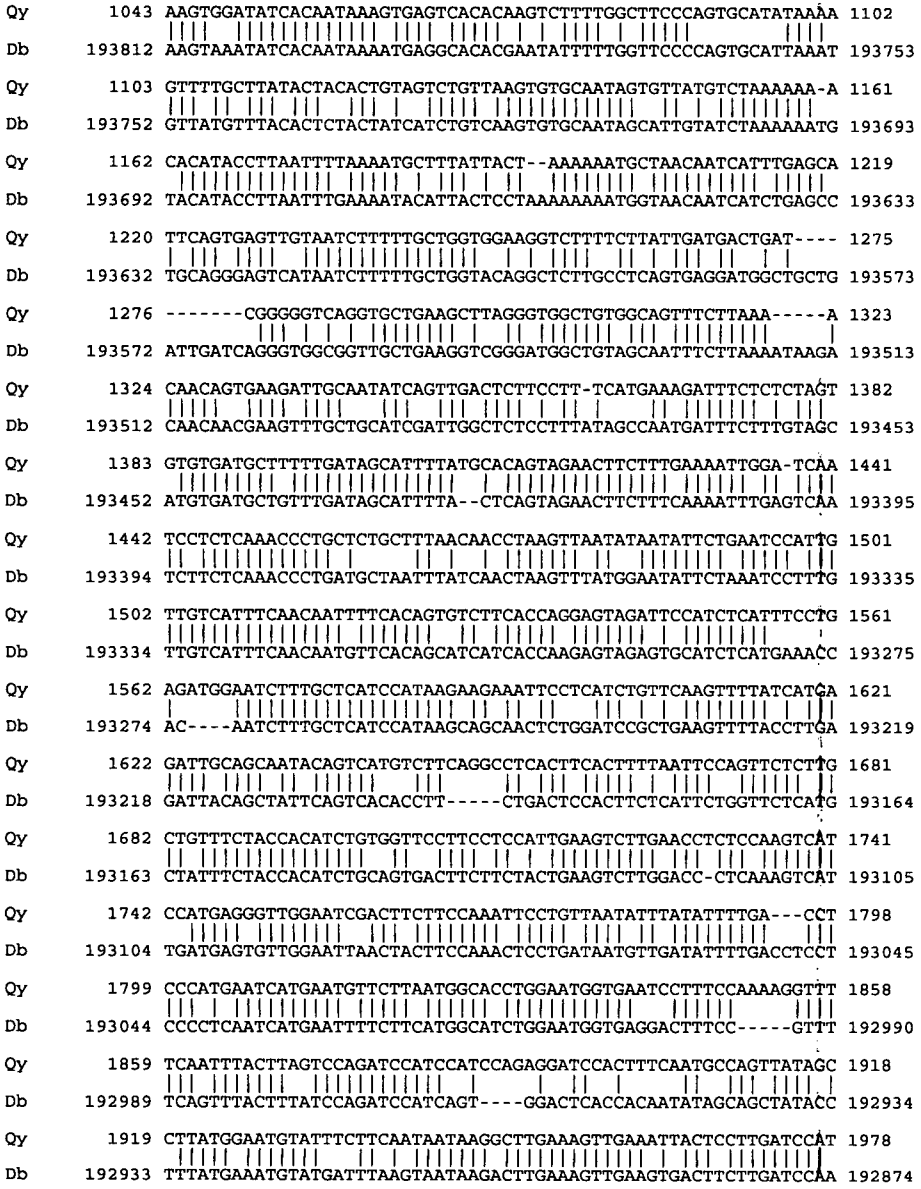
Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
              ||| |||||||  ||||||||||||  |     |  ||  |     ||  ||| |||| |
Db     192989 TCAGTTTACTTTATCCAGATCCATCAGT----GGACTCACCACAATATAGCAGCTATACC 192934

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              |||||  |||||||   ||  | ||||||| |||||||||||||  | |||  |||||||||
Db     192933 TTTATGAAATGTATGATTTAAGTAATAAGACTTGAAAGTTGAAGTGACTTCTTGATCCAA 192874

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000
                |||||  |||  |||||||||
Db     192873 GGACTGCAGAATGGATGTTGTG 192852


RESULT 43
US-09-949-016-14554/c
; Sequence 14554, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14554
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14554

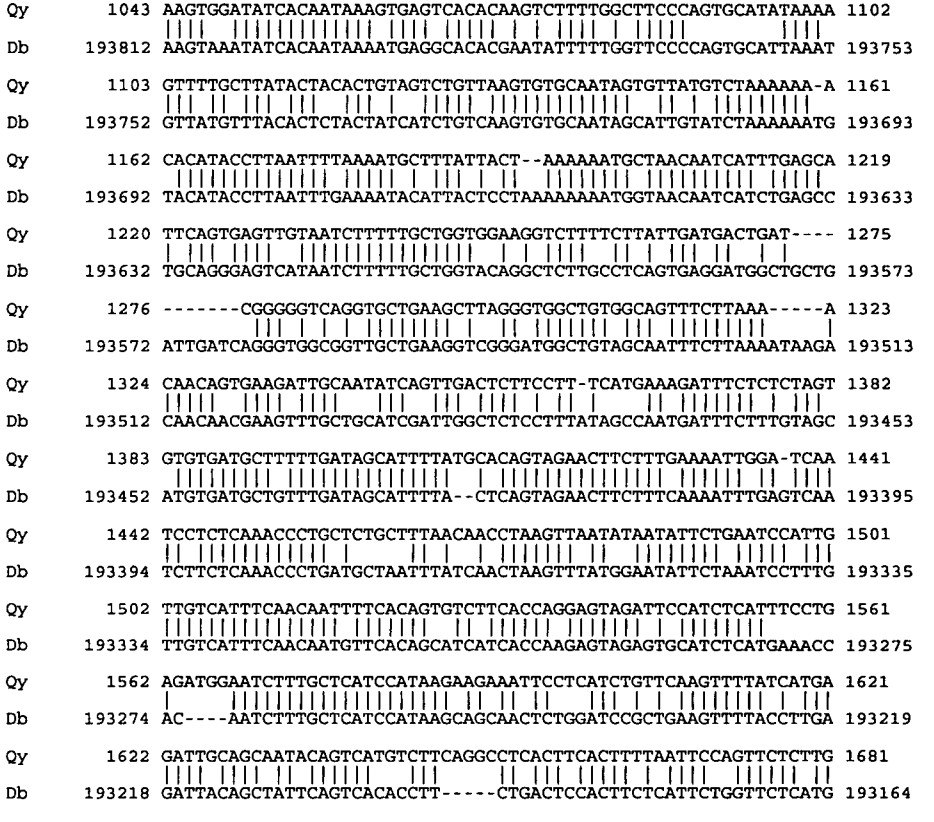
  Query Match             21.3%;  Score 426;  DB 3;  Length 818128;
  Best Local Similarity   74.0%;  Pred. No. 6.8e-78;
  Matches  727;  Conservative    0;  Mismatches  210;  Indels   45;  Gaps   13;
```



RESULT 44
US-09-949-016-14555/c
; Sequence 14555, Application US/09949016

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14555
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14555

Query Match 21.3%; Score 426; DB 3; Length 818128;
Best Local Similarity 74.0%; Pred. No. 6.8e-78;
Matches 727; Conservative 0; Mismatches 210; Indels 45; Gaps 13;



```
Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741
             || ||||||||||||||||  ||  |||| || | |||||||||| ||| ||| |||||||
Db     193163 CTATTTCTACCACATCTGCAGTGACTTCTTCTACTGAAGTCTTGGACC-CTCAAAGTCAT 193105

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCT 1798
              ||||| ||||||||  ||| ||||||||  |||| |||| || |||||||||    |||
Db     193104 TGATGAGTGTTGGAATTAACTACTTCCAAACTCCTGATAATGTTGATATTTTGACCTCCT 193045

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858
              ||| | |||||||||| ||||| |||||| ||||||||||||    ||||||     ||||
Db     193044 CCCCTCAATCATGAATTTTCTTCATGGCATCTGGAATGGTGAGGACTTTCC-----GTTT 192990

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
              ||| |||||||  |||||||||||| |     | ||  |      ||  ||| |||| |
Db     192989 TCAGTTTACTTTATCCAGATCCATCAGT----GGACTCACCACAATATAGCAGCTATACC 192934

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              ||||| |||||||   || | ||||||| |||||||||||| | ||| |||||||||
Db     192933 TTTATGAAATGTATGATTTAAGTAATAAGACTTGAAAGTTGAAGTGACTTCTTGATCCAA 192874

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000
                ||||| ||| |||||||||
Db     192873 GGACTGCAGAATGGATGTTGTG 192852


RESULT 45
US-09-949-016-14556/c
; Sequence 14556, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14556
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14556

  Query Match             21.3%;  Score 426;  DB 3;  Length 818128;
  Best Local Similarity   74.0%;  Pred. No. 6.8e-78;
  Matches  727;  Conservative    0;  Mismatches  210;  Indels   45;  Gaps   13;

Qy       1043 AAGTGGATATCACAATAAAGTGAGTCACACAAGTCTTTTGGCTTCCCAGTGCATATAAAA 1102
              ||||  |||||||||||||| |||| ||||| | | |||| | |||||        ||||
Db     193812 AAGTAAATATCACAATAAAATGAGGCACACGAATATTTTTGGTTCCCCAGTGCATTAAAT 193753

Qy       1103 GTTTTGCTTATACTACACTGTAGTCTGTTAAGTGTGCAATAGTGTTATGTCTAAAAAA-A 1161
              ||| || ||| |||  ||| |  ||||| |||||||||||||  || | |||||||||
Db     193752 GTTATGTTTACACTCTACTATCATCTGTCAAGTGTGCAATAGCATTGTATCTAAAAAATG 193693

Qy       1162 CACATACCTTAATTTTAAAATGCTTTATTACT--AAAAAATGCTAACAATCATTTGAGCA 1219
              |||||||||||||| |||||  | ||| | ||  |||||||| |||||||||| |||||
Db     193692 TACATACCTTAATTTGAAAATACATTACTCCTAAAAAAAATGGTAACAATCATCTGAGCC 193633

Qy       1220 TTCAGTGAGTTGTAATCTTTTTGCTGGTGGAAGGTCTTTTCTTATTGATGACTGAT---- 1275
              | ||| ||||  |||||||||||||||  | | ||||  || | ||| ||  | |
Db     193632 TGCAGGGAGTCATAATCTTTTTGCTGGTACAGGCTCTTGCCTCAGTGAGGATGGCTGCTG 193573

Qy       1276 -------CGGGGGTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAAA-----A 1323
                    ||| | | |||||||| |  || ||||||| ||| |||||||||     |
Db     193572 ATTGATCAGGGTGGCGGTTGCTGAAGGTCGGGATGGCTGTAGCAATTTCTTAAAATAAGA 193513

Qy       1324 CAACAGTGAAGATTGCAATATCAGTTGACTCTTCCTT-TCATGAAAGATTTCTCTCTAGT 1382
              |||||  |||| ||||   |||  ||| |||| | || |    || ||||||| | |||
Db     193512 CAACAACGAAGTTTGCTGCATCGATTGGCTCTCCTTTATAGCCAATGATTTCTTTGTAGC 193453

Qy       1383 GTGTGATGCTTTTTGATAGCATTTTATGCACAGTAGAACTTCTTTGAAAATTGGA-TCAA 1441
              ||||||||| |||||||||||||||  | ||||||||||||||| |||||| || ||||
Db     193452 ATGTGATGCTGTTTGATAGCATTTTA--CTCAGTAGAACTTCTTTCAAAATTTGAGTCAA 193395

Qy       1442 TCCTCTCAAACCCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTCTGAATCCATTG 1501
              || |||||||||||| |      || |  | ||||||| ||  |||||||| ||||| |||
Db     193394 TCTTCTCAAACCCTGATGCTAATTTATCAACTAAGTTTATGGAATATTCTAAATCCTTTG 193335

Qy       1502 TTGTCATTTCAACAATTTTCACAGTGTCTTCACCAGGAGTAGATTCCATCTCATTTCCTG 1561
              |||||||||||||||| |||||||  || |||||| ||||||| | ||||||||
Db     193334 TTGTCATTTCAACAATGTTCACAGCATCATCACCAAGAGTAGAGTGCATCTCATGAAACC 193275

Qy       1562 AGATGGAATCTTTGCTCATCCATAAGAAGAAATTCCTCATCTGTTCAAGTTTTATCATGA 1621
              |    |||||||||||||||||||| || || ||   ||| | | |||||||| | |||
Db     193274 AC----AATCTTTGCTCATCCATAAGCAGCAACTCTGGATCCGCTGAAGTTTTACCTTGA 193219

Qy       1622 GATTGCAGCAATACAGTCATGTCTTCAGGCCTCACTTCACTTTTAATTCCAGTTCTCTTG 1681
              |||| |||| || ||||||   |||      || ||| ||||| | ||||  ||||| ||
Db     193218 GATTACAGCTATTCAGTCACACCTT-----CTGACTCCACTTCTCATTCTGGTTCTCATG 193164

Qy       1682 CTGTTTCTACCACATCTGTGGTTCCTTCCTCCATTGAAGTCTTGAACCTCTCCAAGTCAT 1741
              || |||||||||||||  ||  |||| || | |||||||||| ||| ||| |||||||
Db     193163 CTATTTCTACCACATCTGCAGTGACTTCTTCTACTGAAGTCTTGGACC-CTCAAAGTCAT 193105

Qy       1742 CCATGAGGGTTGGAATCGACTTCTTCCAAATTCCTGTTAATATTTATATTTTGA---CCT 1798
              ||||| ||||||||  ||| ||||||||  |||| |||| || |||||||||    |||
Db     193104 TGATGAGTGTTGGAATTAACTACTTCCAAACTCCTGATAATGTTGATATTTTGACCTCCT 193045

Qy       1799 CCCATGAATCATGAATGTTCTTAATGGCACCTGGAATGGTGAATCCTTTCCAAAAGGTTT 1858
              ||| | |||||||||| ||||| |||||| ||||||||||||    ||||||     ||||
Db     193044 CCCCTCAATCATGAATTTTCTTCATGGCATCTGGAATGGTGAGGACTTTCC-----GTTT 192990

Qy       1859 TCAATTTACTTAGTCCAGATCCATCCATCCAGAGGATCCACTTTCAATGCCAGTTATAGC 1918
              ||| |||||||  |||||||||||| |     | ||  |      ||  ||| |||| |
Db     192989 TCAGTTTACTTTATCCAGATCCATCAGT----GGACTCACCACAATATAGCAGCTATACC 192934

Qy       1919 CTTATGGAATGTATTTCTTCAATAATAAGGCTTGAAAGTTGAAATTACTCCTTGATCCAT 1978
              ||||| |||||||   || | ||||||| |||||||||||| | ||| |||||||||
Db     192933 TTTATGAAATGTATGATTTAAGTAATAAGACTTGAAAGTTGAAGTGACTTCTTGATCCAA 192874

Qy       1979 TTTCTGCAAAATAGATGTTGTG 2000
                ||||| ||| |||||||||
Db     192873 GGACTGCAGAATGGATGTTGTG 192852


Search completed: January 22, 2006, 21:29:36
Job time : 289.102 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:10:03 ; Search time 5419.54 Seconds
(without alignments)
17274.699 Million cell updates/sec

Title: US-09-728-552A-3_COPY_19000_21000
Perfect score: 2001
Sequence: 1 aatgtttggatttcaaaggt.....cagggacatgggtgcatagc 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	207.6	10.4	415	1	AI609972	AI609972 tt78c10.x
c 2	206.4	10.3	490	9	AQ426532	AQ426532 CITBI-E1-
3	205.2	10.3	463	9	AQ535344	AQ535344 RPCI-11-3
c 4	204.2	10.2	412	2	BE062478	BE062478 QV4-BT025
5	202	10.1	590	9	AQ421062	AQ421062 RPCI-11-1
c 6	201.6	10.1	390	5	BU588888	BU588888 AGENCOURT
7	200.8	10.0	711	11	CR959688	CR959688 Homo sapi
8	200	10.0	533	7	CN274852	CN274852 170006000
c 9	200	10.0	687	10	AG085195	AG085195 Pan trogl
c 10	199.6	10.0	660	9	AQ376970	AQ376970 RPCI11-15
11	199.4	10.0	461	9	AQ112243	AQ112243 CIT-HSP-2
12	199.2	10.0	655	6	CA439825	CA439825 UI-H-DI0-
c 13	199	9.9	694	9	BZ603010	BZ603010 WHAAX35TF
14	198.8	9.9	467	5	BU735183	BU735183 UI-E-DW0-
c 15	198.8	9.9	591	9	AQ469410	AQ469410 CITBI-E1-
16	198.6	9.9	439	1	AA515728	AA515728 ng70f04.s
17	198.4	9.9	455	1	AA904211	AA904211 od88e02.s
c 18	198.4	9.9	521	1	AW970571	AW970571 EST382652
c 19	198.4	9.9	616	9	AQ554309	AQ554309 RPCI-11-4
20	198.4	9.9	990	3	BM803650	BM803650 AGENCOURT
c 21	198.2	9.9	711	9	AQ415030	AQ415030 RPCI-11-2
22	198	9.9	441	1	AA169245	AA169245 zp19d03.s
23	198	9.9	751	9	AQ788367	AQ788367 HS_3135_A
c 24	197.8	9.9	322	5	BU603903	BU603903 AGENCOURT
c 25	197.8	9.9	324	5	BU588173	BU588173 AGENCOURT
c 26	197.8	9.9	757	9	AQ528478	AQ528478 RPCI-11-3
c 27	197.8	9.9	896	9	AQ738849	AQ738849 HS_5386_B
28	197.6	9.9	493	5	BX486751	BX486751 DKFZp686E
c 29	197.6	9.9	607	9	BZ603705	BZ603705 WHADP21TR
30	197.6	9.9	730	5	BU617273	BU617273 UI-H-DF0-
31	197.6	9.9	741	6	CA428248	CA428248 UI-H-DF0-
c 32	197.4	9.9	690	9	AQ415537	AQ415537 RPCI-11-1
c 33	197.4	9.9	720	10	CL423123	CL423123 RP11-429A
c 34	197.2	9.9	660	9	AQ471471	AQ471471 CITBI-E1-
35	196.8	9.8	410	1	AA832175	AA832175 oc91c04.s
36	196.8	9.8	2978	4	CR861212	CR861212 Pongo pyg
c 37	196.6	9.8	634	6	CB154854	CB154854 K-EST0212
38	196.6	9.8	732	9	AQ506289	AQ506289 RPCI-11-2
c 39	196.4	9.8	669	10	AG182560	AG182560 Pan trogl
c 40	196.4	9.8	751	9	BZ612069	BZ612069 WHADP28TR
41	196.2	9.8	3110	4	HSM801883	AL136917 Homo sapi
42	196	9.8	510	9	AQ885682	AQ885682 HS_5525_A
43	196	9.8	724	9	AQ379603	AQ379603 RPCI11-16
44	195.8	9.8	600	9	AQ075668	AQ075668 CIT-HSP-2
45	195.8	9.8	3015	4	CR859293	CR859293 Pongo pyg
46	195.8	9.8	3775	4	CR627381	CR627381 Homo sapi
c 47	195.6	9.8	324	2	BG029528	BG029528 602297087
c 48	195.6	9.8	644	10	AG057657	AG057657 Pan trogl
49	195.4	9.8	633	5	BX480642	BX480642 DKFZp686I
50	195.4	9.8	1058	3	BM554734	BM554734 AGENCOURT
c 51	195.4	9.8	2820	4	AF318360	AF318360 Homo sapi
c 52	195.2	9.8	643	2	BG714390	BG714390 602669912
c 53	195	9.7	645	7	CK001378	CK001378 AGENCOURT
54	195	9.7	857	6	CB997964	CB997964 AGENCOURT
c 55	195	9.7	929	6	CD300528	CD300528 AGENCOURT
c 56	195	9.7	3744	4	HSM807621	BX647475 Homo sapi
c 57	195	9.7	4207	4	CR858093	CR858093 Pongo pyg
c 58	194.8	9.7	362	5	BU569947	BU569947 AGENCOURT
c 59	194.8	9.7	593	9	AQ078004	AQ078004 CIT-HSP-2
c 60	194.8	9.7	679	10	AG123655	AG123655 Pan trogl
c 61	194.8	9.7	719	9	AQ475181	AQ475181 CITBI-E1-
62	194.6	9.7	518	5	BU659257	BU659257 cl45e06.z
c 63	194.6	9.7	544	1	AL037856	AL037856 DKFZp564I
64	194.6	9.7	551	9	AQ429304	AQ429304 CITBI-E1-
65	194.6	9.7	553	9	AQ481091	AQ481091 RPCI-11-2
c 66	194.6	9.7	592	5	BU579561	BU579561 im86e10.y
67	194.6	9.7	602	1	AL138330	AL138330 DKFZp547O
68	194.6	9.7	649	11	CR960737	CR960737 Homo sapi
69	194.6	9.7	788	8	CX782424	CX782424 HESC3_14
70	194.6	9.7	919	5	BU844893	BU844893 AGENCOURT
c 71	194.4	9.7	508	9	AQ321648	AQ321648 RPCI11-93
c 72	194.4	9.7	550	1	AV756491	AV756491 AV756491
c 73	194.4	9.7	714	10	AG165194	AG165194 Pan trogl
74	194.4	9.7	719	2	BG221939	BG221939 RST41756
75	194.4	9.7	730	9	AQ489977	AQ489977 RPCI-11-2
c 76	194.4	9.7	808	9	BZ611126	BZ611126 WHADL87TF
77	194.2	9.7	417	9	AQ388549	AQ388549 RPCI11-14
c 78	194.2	9.7	541	9	BZ604634	BZ604634 WHAAP93TR
c 79	194.2	9.7	647	10	AG039308	AG039308 Pan trogl
c 80	194.2	9.7	840	4	BC019333	BC019333 Homo sapi
81	194.2	9.7	917	6	CD558585	CD558585 AGENCOURT
c 82	194.2	9.7	1987	4	AF289581	AF289581 Homo sapi
c 83	194	9.7	320	1	AI306232	AI306232 qw80b05.x
c 84	194	9.7	535	9	AQ310823	AQ310823 CIT-HSP-2
c 85	194	9.7	2234	4	AY203938	AY203938 Homo sapi
86	193.8	9.7	393	1	AI491765	AI491765 tn94g03.x
87	193.8	9.7	503	9	AQ800087	AQ800087 HS_5250_A
c 88	193.8	9.7	553	9	AQ387600	AQ387600 RPCI11-14
c 89	193.8	9.7	581	9	AQ347610	AQ347610 RPCI11-12
90	193.8	9.7	702	6	CA428680	CA428680 UI-H-DF0-
91	193.8	9.7	777	2	BG193279	BG193279 RST12407
92	193.6	9.7	508	5	BX646464	BX646464 DKFZp781C
c 93	193.6	9.7	577	9	AQ265389	AQ265389 CITBI-E1-
94	193.6	9.7	749	6	CB962143	CB962143 AGENCOURT
c 95	193.4	9.7	543	9	AQ079748	AQ079748 CIT-HSP-2

```
 c  96   193.4    9.7    575  9  AQ587429                  AQ587429 CITBI-E1-
 c  97   193.4    9.7    638  6  CD673070                  CD673070 fg20b08.y
    98   193.4    9.7    677  10  AG094571                  AG094571 Pan trogl
 c  99   193.4    9.7    691  9  AQ899804                  AQ899804 HS_2013_A
 c 100   193.4    9.7    765  6  CD516602                  CD516602 AGENCOURT
```

ALIGNMENTS

```
RESULT 1
AI609972/c
LOCUS       AI609972                415 bp    mRNA    linear   EST 21-APR-1999
DEFINITION  tt78c10.x1 NCI_CGAP_HSC3 Homo sapiens cDNA clone IMAGE:2246898 3'
            similar to contains Alu repetitive element;contains element MER9
            repetitive element ;, mRNA sequence.
ACCESSION   AI609972
VERSION     AI609972.1  GI:4619139
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2246898"
                     /tissue_type="CD34+, T negative, patient with chronic
                     myelogenous leukemia"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_HSC3"
                     /note="Organ: bone marrow; Vector: pAMP1; mRNA made from
                     lymphoid tissue, cDNA made by oligo-dT  priming.
                     Directionally cloned.  Size-selected on agarose  gel,
                     average insert size 500 bp.  Primary library,
                     non-amplified.   cDNA Library Preparation: David B.
                     Krizman, Ph.D. Reference:  Krizman et al. (1996) Cancer
                     Research 56:5380-5383."
ORIGIN

  Query Match           10.4%;  Score 207.6;  DB 1;  Length 415;
  Best Local Similarity   84.8%;  Pred. No. 1.9e-24;
  Matches  280;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    4;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| |||||||||| |||||||||| || | ||| |||||||||||| ||
Db        400 TTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCGCTGTTTCATCCAGGCTGGAGTGCA 341

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| |||| |||||||||||||||||||| |||||||||||||||| |||||||||
Db        340 GTGG-TGTGATCTCGGCTCACTGCAACCTC-GCCTCCCAGGTTCAAGCAATTCTCCTGCC 283

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTA 1391
             |||||||||| | |||||||||||||||||||||||  |  ||   | ||||||||||||
Db        282 TCAGCTTCCCAAGTAGCTGGGATTACAGGCGCATGCCACCACACCCGGCTAATTTTTGTA 223

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             |||||||||||||||| |||| |||||||| ||||||| | ||||||||||||| |||
Db        222 TTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCA 163

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
             ||| |||| |||||||| |||||||||| |||||||||||||| ||| |||| ||||
Db        162 GGTAATCCGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCAAGAGACACCATGCC 103

Qy       1512 CAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             |||||| | || ||  ||  |  | || ||
Db        102 CAGCCTCTCTTTTCTTATAATCAACTTCTT 73


RESULT 2
AQ426532/c
LOCUS       AQ426532                490 bp    DNA     linear   GSS 24-MAR-1999
DEFINITION  CITBI-E1-2574H2.TR CITBI-E1 Homo sapiens genomic clone 2574H2,
            genomic survey sequence.
ACCESSION   AQ426532
VERSION     AQ426532.1  GI:4499300
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2574H2"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
ORIGIN

  Query Match           10.3%;  Score 206.4;  DB 9;  Length 490;
  Best Local Similarity   78.0%;  Pred. No. 2.9e-24;
  Matches  287;  Conservative    0;  Mismatches   76;  Indels    5;  Gaps    3;

Qy       1155 AAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAG 1214
             || || |  |||||||    || || |   |  |||   |   | || |   |  |
Db        433 AAGGAAAATCCATTTCCAGATTAAAGTACATTTAAGAGAAAAACAAATAAATGCAGTGTG 374

Qy       1215 ATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTAC 1274
             |  |||  || ||  |||| |||| |||||||||||||| ||||||||||||||||||| |
Db        373 GTTATTTATTTATTTATTTTATTGAGACAGAGTCTCACTCTGTCACCCAGGCTGGAGTGC 314

Qy       1275 AGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG- 1333
              |||||   | |||||||||||||||||||| ||||||| ||||||||||||||||||
```

```
Db        313 TGTGGC-ACGATCTCGGCTCACTGCAACCTCCGCCTCCCCGGTTCAAGCGATTCTCCTGC 255

Qy       1334 -TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGT 1390
              ||||||||||||| | ||||||||||||||||||  ||   |  ||    ||||||||||
Db        254 CTCAGCTTCCCGAGTTGCTGGGATTACAGGCGCCCGCCACCACCCCCAGCTAATTTTTGT 195

Qy       1391 ATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTC 1450
              |||||||||||||||  |  |||| |||||||| |||||||  | || ||||||||||| ||
Db        194 ATTTTTAGTAGAGACGGGCTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTC 135

Qy       1451 AGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGC 1510
              |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||  ||
Db        134 AGGTGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGC 75

Qy       1511 CCAGCCTG 1518
                | |||||
Db         74 TCGGCCTG 67


RESULT 3
AQ535344
LOCUS       AQ535344                463 bp    DNA     linear   GSS 18-MAY-1999
DEFINITION  RPCI-11-356I15.TV RPCI-11 Homo sapiens genomic clone
            RPCI-11-356I15, genomic survey sequence.
ACCESSION   AQ535344
VERSION     AQ535344.1  GI:4847034
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-356I15.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7636526"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-356I15"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             10.3%;  Score 205.2;  DB 9;  Length 463;
  Best Local Similarity   73.5%;  Pred. No. 4.7e-24;
  Matches  302;  Conservative    0;  Mismatches  104;  Indels    5;  Gaps    3;

Qy       1173 TGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTT 1232
              |||||    | ||    ||    | |   ||| | ||  | |  | | || |||||  |||
Db         33 TGTTGGTTATATTTCTGGACTCCTATTTGTATCTATTGACTGTCCATCTTTTTTTTTTTT 92

Qy       1233 TTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGC 1292
              ||||||| |||||||||| || |||  |||||||||||||| |||||||  ||| | || ||
Db         93 TTTTTGAGACAGAGTCTCCCTCTGTTGCCCAGGCTGGAGTGCAGTGGC-GTGATTTCAGC 151

Qy       1293 TCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGC 1350
              | |||| ||||||||||||||||||||||||||||||| ||||||  ||||| |||| | | ||
Db        152 TTACTGAAACCTCTGCCTCCCAGGTTCAAGCGATCCTCCTGCCTCAGCCTCCCAAGTCGC 211

Qy       1351 TGGGATTACAGG--CGCATGCACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGA 1408
              |||||||| |||  | |||    |    |  | ||||||||||||||||||||||||||||
Db        212 TGGGATTATAGGCACACATCACCATGCCCGACTAATTTTTGTATTTTTAGTAGAGACAGG 271

Qy       1409 GTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTC 1468
              ||||| |||||||| ||||||| | |||||||||||||| ||||||||||||| || ||||
Db        272 GTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCATCCTCCTC 331

Qy       1469 AGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGA 1528
              ||||||||||  ||||||||||||||||| |||| | |||  |||  | | |
Db        332 GGCCTCCCAAAGTGCTGGGATTACAGGCATGAACCACTGCGCCTGGCCACTGTCCATCNT 391

Qy       1529 TCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTCATTTTGGA 1579
              | ||| |   | | |  ||      |||| ||||| ||   ||   ||||
Db        392 TTTTGTTTGGTTGTTTTGAGATGGAGTTTCGATCTTGTTGCCCAGGCTGGA 442


RESULT 4
BE062478/c
LOCUS       BE062478                412 bp    mRNA    linear   EST 09-JUN-2000
DEFINITION  QV4-BT0257-261099-011-b01 BT0257 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE062478
VERSION     BE062478.1  GI:8407128
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV4-BT0257-261
            099-011-b01&t3=1999-10-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 6
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .412
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="BT0257"
```

/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 10.2%; Score 204.2; DB 2; Length 412;
Best Local Similarity 81.3%; Pred. No. 7e-24;
Matches 274; Conservative 0; Mismatches 58; Indels 5; Gaps 3;

```
Qy       1189 AGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTC 1248
              || ||| | | | |   ||||||    | || || ||||| | |||||||| ||||||||
Db        401 AGTAGTTTATCACTTGTTTTTCTTTTCTTTTCTTTTTTTTTTCTTTTTTGAGACAGAGTC 342

Qy       1249 TCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGC 1308
              ||||| |||  |||||||||||| |||||||  ||||||||||||||||||||||||| ||
Db        341 TCACTCTGTTGCCCAGGCTGGAGTGCAGTGGC-ATGGTCTCGGCTCACTGCAACCTCCGC 283

Qy       1309 CTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCA 1366
              ||||| ||| |||| ||||||||||  ||||| ||| || || |||||| ||||||||
Db        282 CTCCCGGGTCCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTACCTGGGACTACAGGCGTG 223

Qy       1367 TGC--ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGAC 1424
              |||   |  |    |||||||||||||||||||||||||||||| ||||| ||| |||| |
Db        222 TGCCACCATGCCCAGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCACGTTGGC 163

Qy       1425 CAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACT 1484
              |||||| | || |||||||||| |||||||||||||| ||||  |||||||||  ||
Db        162 CAGGCTGGTCTGGAACTCCTGACCTCAGGTGATCCACCCGCCTCCACCTCCCAAAGTGCT 103

Qy       1485 GGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
              ||||||||||||||||||||| ||||| ||| ||||
Db        102 GGGATTACAGGCATGAGCCACCATGCCCGGCCAGTTT 66
```

RESULT 5
AQ421062
LOCUS AQ421062 590 bp DNA linear GSS 23-MAR-1999
DEFINITION RPCI-11-188O10.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-188O10, genomic survey sequence.
ACCESSION AQ421062
VERSION AQ421062.1 GI:4478786
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 590)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-188O10.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.

FEATURES Location/Qualifiers
source 1. .590
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7572153"
/db_xref="taxon:9606"
/clone="RPCI-11-188O10"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"

ORIGIN

Query Match 10.1%; Score 202; DB 9; Length 590;
Best Local Similarity 84.1%; Pred. No. 1.5e-23;
Matches 264; Conservative 0; Mismatches 45; Indels 5; Gaps 3;

```
Qy       1207 TTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGC 1266
              | | | |  |  |||| ||| ||||||||| || |||||||||| || |||| |||||||
Db        146 TGTTTTACTTTCTTTTCTTTCTATTTTTTTGGAGACAGAGTCTCGCTCTGTCGCCCAGGC 205

Qy       1267 TGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGAT 1326
              || ||| ||||||| |   ||||||||||||||||||||  ||||||  ||||||||||||
Db        206 TGAAGTGCAGTGGC-GCAATCTCGGCTCACTGCAACCTCCACCTCCCGGGTTCAAGCGAT 264

Qy       1327 TCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTA 1382
              |||||||  |||||||||| || |||||||||||||||||||| ||  |  ||  |  |||
Db        265 TCTCCTGCCTCAGCTTCCTGAGTAGCTGGGATTACAGGCGCACGCCACCACACCCAGCTA 324

Qy       1383 ATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTC 1442
              |||||||||||||||||||||||||||||||| ||||||||  ||||||  | |||||||||
Db        325 ATTTTTGTATTTTTAGTAGAGACAGAGTTTCACCATGTTGGTCAGGCTGGTCTTGAACTC 384

Qy       1443 CTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGC 1502
              |||||  ||||||  |||||||  ||||||||||||||||  ||||||||||||||| |||||
Db        385 CTGACCTCAGGTTATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGC 444

Qy       1503 CACCGTGCCCAGCC 1516
              | |||||||||| |
Db        445 CGCCGTGCCCAGAC 458
```

RESULT 6
BU588888/c
LOCUS BU588888 390 bp mRNA linear EST 20-SEP-2002
DEFINITION AGENCOURT_8910326 NIH_MGC_141 Homo sapiens cDNA clone IMAGE:6387393 5', mRNA sequence.
ACCESSION BU588888
VERSION BU588888.1 GI:23240250
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 390)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2587 row: c column: 10
High quality sequence stop: 389.
FEATURES Location/Qualifiers
source 1. .390

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6387393"
                     /tissue_type="mixed (pool of 40 RNAs)"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_141"
                     /note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc);
                     Site_2: SfiI (ggccgcctcggcc); Double-stranded cDNA was
                     prepared from a pool of 40 cell line polyA+ RNAs (bladder
                     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
                     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
                     kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
                     ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
                     gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
                     used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.2-0.5
                     kb size fraction (other fractions present in NIH_MGC_142).
                     Library created in the laboratory of M. Brownstein (NIMH,
                     NIH). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             10.1%;  Score 201.6;  DB 5;  Length 390;
  Best Local Similarity   77.2%;  Pred. No. 1.9e-23;
  Matches  284;  Conservative    0;  Mismatches   79;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| |||||||||  || |||||||| || |||| ||||||||||||  ||
Db        377 TTTTTTTTTTTTTTTTTTTTTGGGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGGGCA 318

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             ||||| | | ||||||||||||||||| ||| ||||||| |||||| || |||||||||
Db        317 GTGGC-GCGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCC 259

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTA 1391
             ||||| |||||| |||||||||||||||| |  ||  |  ||  |  ||||||||||||
Db        258 TCAGCCTCCCGAGTAGCTGGGATTACAGGCACCCGCCACCACAGCCAGCTAATTTTTGTA 199

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             |||||||||||||| | ||||| ||||||||  |||||| | |||| |||||||||  |||
Db        198 TTTTTAGTAGAGACGGGGTTTCACCATGTTGGTCAGGCTGGTCTTGCACTCCTGACCTCA 139

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
             |||||| ||||| ||||| ||||||||||  |||||||||||||| |||||||||  |||
Db        138 GGTGATTCACCCGCCTCAACCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCC 79

Qy       1512 CAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTC 1571
             |||||||   ||   | | | |||   ||  | ||  | |   |||| ||  |
Db          78 CAGCCTGAAAACTTTTTTTTTTTTTTTTGAGATGGTGTCTCGCTCTTGCTCTTGTTGCCC 19

Qy       1572 ATTTTGGA 1579
             |   ||||
Db          18 AGGCTGGA 11


RESULT 7
CR959688
LOCUS       CR959688                711 bp    DNA     linear   GSS 10-JUN-2005
DEFINITION  Homo sapiens BAC end sequence of RZPDB737H082009D from genomic
            library (orig. Pieter J. de Jong library RPCI-11), genomic survey
            sequence.
ACCESSION   CR959688
VERSION     CR959688.1  GI:66957175
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Schwarz,F., Neubert,P., Schneider,D., Peters,M. and Korn,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2005) RZPD Deutsches Ressourcenzentrum fuer
            Genomforschung GmbH, Im Neuenheimer Feld 515, D-69120 Heidelberg,
            Germany
COMMENT     RZPD; RZPDB737H082009D; RP11-17P16;
            derived from Pieter J. de Jong library RPCI-11;
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPDB737H082009D
            RZPDLIB; (Human Genomic Set - RZPD 1.0) RZPD LIB No.737
            http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=737
            http://www.rzpd.de/products/genomicset/
            Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 100
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available from RZPD;
            Contact RZPD (customer.service@rzpd.de) for further information.
            Clone distribution: http://www.rzpd.de/products/genomicset/
            Seq-primer: T7 (TAA-TAC-GAC-TCA-CTA-TAG-GG)
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RZPDB737H082009D"
                     /sex="Male"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; RPCI-11 Human Male BAC Library"
ORIGIN

  Query Match             10.0%;  Score 200.8;  DB 11;  Length 711;
  Best Local Similarity   67.3%;  Pred. No. 2.2e-23;
  Matches  327;  Conservative    0;  Mismatches  154;  Indels    5;  Gaps    3;

Qy       1094 TCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTTAGCCATGTGTGAC 1153
             | ||  ||| |  || | ||   | | |  | ||| |     |||||    | ||  |
Db        179 TAATTATCCTTTTAATTTCTGTGGTATCAGTTGTAATGTCTCCTTTTAAATCTCTGATAG 238

Qy       1154 AAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCA 1213
              | | |   |    |||  | ||         | ||  |||  ||       | | || |
Db        239 TATATATTTGAGCCTTCTTTTTTTCTTAGTCTGAGTAAAGGTTTGTTGATTTTCTTTTTT 298

Qy       1214 GATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTA 1273
              | || || ||||| || |||||||  |  |||||||  | |||||||||  |||||||
Db        299 CTTTTTCTTTTTTTTTTTTATTTTTGAGATGGAGTCTCAGTCTGTCACCCAGACTGGAGTG 358

Qy       1274 CAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG 1333
              |||||||     |||| ||||||||||||||  ||||||  |||||||||||||||||||
Db        359 CAGTGGC-ACAATCTCAGCTCACTGCAACCTCCACCTCCCGGGTTCAAGCGATTCTCCTG 417

Qy       1334 --TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACC--ATGCCTAATTTTTG 1389
                |||||  |||| | || ||||||||||||| |  |||  |  |   || || |||||||
Db        418 CCTCAGCCTCCCAAGTACCTGGGATTACAGGTGTGTGCCACCACGGCTGGCTCATTTTTG 477

Qy       1390 TATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTT 1449
              |||||||||||||||| ||||| |||||||| |||||||  | || ||||||||||  |
Db        478 TATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCT 537

Qy       1450 CAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTG 1509
              || ||||||||||  |||| | |||||||  | |  ||||||||||||  |||||||||  ||
Db        538 CAAGTGATCCACCTGCCTCGGACTCCCAATGTATTGGGATTACAGGTGTGAGCCACCATG 597

Qy       1510 CCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTAT 1569
               ||  || | |||| ||| |     | ||  ||     || ||||     | | | |   | |
Db        598 CCTGGCTTATTTTTTCAAAAAACTTTTTGTTTNAATAATCCTTTGNATTTTTTTAAATGT 657

Qy       1570 TCATTT 1575
              ||||
Db        658 AAATTT 663
```

```
RESULT 8
CN274852
LOCUS       CN274852                533 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000600044196 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN274852
VERSION     CN274852.1  GI:47291266
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
   PUBMED   15146197
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 533   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .533
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cell, retinoic acid and
                     mitogen-treated hES cell line H7"
                     /clone_lib="GRN_PRENEU"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from hES cell line H7 (p29) maintained in feeder-free
                     conditions. Embryoid bodies were generated in the presence
                     of all-trans retinoic and mitogens."
ORIGIN

  Query Match             10.0%;  Score 200;  DB 7;  Length 533;
  Best Local Similarity   80.7%;  Pred. No. 3.3e-23;
  Matches  271;  Conservative    0;  Mismatches   60;  Indels    5;  Gaps    3;

Qy       1190 GAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCT 1249
                ||  | ||   ||||  ||| | |    ||  | | ||||||| |||||| |||||||||
Db         10 GAGTTTTTAAGTTATTATTTTTAATTAATTAATTTATTTATTTATTTTGAGACAGAGTCT 69

Qy       1250 CACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCC 1309
                |||| |||| |||||||||||||||||||| |||| ||| ||||||||||||||| |||
Db         70 CACTCTGTCGCCCAGGCTGGAGTACAGTGG-TGTGATCTTGGCTCACTGCAACCTCCGCC 128

Qy       1310 TCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCAT 1367
                ||||||||||||||||||  ||  | ||| |||| | |||||||| |||||||||   |
Db        129 CCCCAGGTTCAAGCGATTCTTGTGCCTGAGCCTCCCTAGTAGCTGGAATTACAGGCAACT 188

Qy       1368 GCACCACCA--TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACC 1425
                ||  |  ||  || ||||||||||||||||||||||||||||| ||||| |||||||| ||
Db        189 GCCACTGCACCTGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCC 248

Qy       1426 AGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTG 1485
                |||||  | |||  | |||||||| |||||||||||||||||||||| ||||||||||  |||
Db        249 AGGCTGGTCTTCAGCTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTG 308

Qy       1486 GGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
                ||||||||||  |||||||||  ||||  |||   ||
Db        309 GGATTACAGGTGTGAGCCACCATGCCTGGCCAAGTT 344


RESULT 9
```

```
AG085195/c
LOCUS       AG085195                687 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-083C18.F, genomic survey sequence.
ACCESSION   AG085195
VERSION     AG085195.1  GI:16636997
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 687)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-083C18.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             10.0%;  Score 200;  DB 10;  Length 687;
  Best Local Similarity   84.0%;  Pred. No. 3.1e-23;
  Matches  262;  Conservative    0;  Mismatches   45;  Indels    5;  Gaps    3;

Qy       1215 ATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTAC 1274
                || | ||| |  || ||||||||||| |||||||||||| |||||||||||||||||| |
Db        451 ATTTATTTTTACTTGTTTTTTTTGAGACAGAGTCTCACTCTGTCACCCAGGCTGGAGTGC 392

Qy       1275 AGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG- 1333
                |||||| ||| |||||||||||||||||||||||||||||||||||| ||| |||| |||
Db        391 AGTGGC-GTGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCGAGCAATTCCCCTGC 333

Qy       1334 -TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGT 1390
                || || ||  || ||||||||| ||||| |||||||  |  ||  |  ||||||||||
Db        332 CTCGGCCTCTTGAGTAGCTGGGACTACAGACGCATGCCACCACACCCAGCTAATTTTTGT 273

Qy       1391 ATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTC 1450
                |||||||||||||||| ||||| |||||||| |||||||| | ||  ||||||||| ||
Db        272 ATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTC 213

Qy       1451 AGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGC 1510
                ||||||||||||||||  | |||||||||  |||||||||||||| |||||||||   |
Db        212 AGGTGATCCACCCACCTTGGGCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACAC 153

Qy       1511 CCAGCCTGTTTT 1522
                || |||| ||||
Db        152 CCGGCCTATTTT 141
```

```
RESULT 10
AQ376970/c
LOCUS       AQ376970                660 bp    DNA     linear   GSS 20-MAY-1999
DEFINITION  RPCI11-151K23.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-151K23,
            genomic survey sequence.
ACCESSION   AQ376970
VERSION     AQ376970.1  GI:4347993
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-151K23.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7557862"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-151K23"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             10.0%;  Score 199.6;  DB 9;  Length 660;
  Best Local Similarity   80.8%;  Pred. No. 3.6e-23;
  Matches  270;  Conservative    0;  Mismatches   59;  Indels    5;  Gaps    3;

Qy       1194 TATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACT 1253
              ||||| || |  |  | ||| | |||||| ||||| |||||||||| |  ||||||||||
Db        372 TATTTTATAAAATGATTTCATAACTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACT 313

Qy       1254 TTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCC 1313
              | ||||||||||||||  || ||| ||   ||||||||||||||||||||| |||||||
Db        312 CTATCACCCAGGCTGGAGGGCAATGG-TGCAATCTCGGCTCACTGCAACCTCCGCCTCCC 254

Qy       1314 AGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCAC 1371
              |||||||| ||||||||||  ||||| |||| | |||||||||||| |||| |||||
Db        253 GGGTTCAAGTGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTAGAGGCACATGCCA 194

Qy       1372 CACCA--TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGC 1429
              |  ||  || ||| |||||||||||||||||||| | | ||||||||||||   ||||||
Db        193 CGGCATCTGGCTATTTTTTGTATTTTTAGTAGAGGCGGGGTTTCGCCATGTTAGCCAGGC 134

Qy       1430 TTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGAT 1489
              | |  |||||||||||| ||||||||||  ||  |||| ||||||||||  |||||||
Db        133 TGGTTTTGAACTCCTGACGTCAGGTGATCGGCCGGCCTCGGCCTCCCAAAGTGCTGGGAT 74

Qy       1490 TACAGGCATGAGCCACCGTGCCCAGCCTGTTTTC 1523
              ||||||  ||| |||| |||||  ||||| ||||
Db          73 TACAGGGGTGAACCACTGTGCCTGGCCTGATTTC 40


RESULT 11
AQ112243
LOCUS       AQ112243                461 bp    DNA     linear   GSS 29-AUG-1998
DEFINITION  CIT-HSP-2373C2.TR CIT-HSP Homo sapiens genomic clone 2373C2,
            genomic survey sequence.
ACCESSION   AQ112243
VERSION     AQ112243.1  GI:3484403
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2373C2.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2373C2"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             10.0%;  Score 199.4;  DB 9;  Length 461;
  Best Local Similarity   74.8%;  Pred. No. 4.3e-23;
  Matches  303;  Conservative    0;  Mismatches   96;  Indels    6;  Gaps    4;

Qy       1152 ACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCT 1211
              ||  |||      |  ||| | || | | |||||||| || | |  ||  ||   ||||||
Db         11 ACTCAAGCTTATCAGTTTTATTGATCACATTGTTCATAGTATTTCTTTCTTTCTTTTTCT 70

Qy       1212 CAGATCTTTTTATTTTTATTTTTTTT-GAAACAGAGTCTCACTTTGTCACCCAGGCTGGA 1270
              |||||||| ||||| |||||||| || |||||||||| | |||| |||||| |||
Db         71 TTTTTCTTTTTTTTTTTTTTTTTTTTCGAGACAGAGTCTCAGTCTGTCGCCCAGGCCGGA 130

Qy       1271 GTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTC 1330
              || |||||| || | ||| |||||||||||||||| |||||| |||| ||||||||||||
Db        131 GTGCAGTGG-TGCGATCTTGGCTCACTGCAACCTCCGCCTCCTAGGTACAAGCGATTCTC 189

Qy       1331 CTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTT 1386
              |||  ||||| |||| | |||||| | ||||||| ||  ||   |   | || |||||||
Db        190 CTGCTTCAGCCTCCCAAGTAGCTGAGGTTACAGGTGCTCGCCACCATGCCTGACTAATTT 249

Qy       1387 TTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGA 1446
```

```
                ||||||||||||||||| || | ||||| |||||||| ||||||| | ||||||||| |||
Db        250 TTGTATTTTTAGTAGAAACGGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCTTGA 309

Qy       1447 CTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACC 1506
                | |||||||||| ||  |||  |||||||||  ||||||||||   ||||||||
Db        310 CCTCAGGTGATCCGCCTGCCTGGACCTCCCAAAGTGCTGGGATTACAGATGTGAGCCACT 369

Qy       1507 GTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCT 1551
                || ||| ||||||||       | || ||||    | | ||  ||
Db        370 GTACCCGGCCTGTTTATAGTATTTCTTTATTATCCTTTTAATGCT 414


RESULT 12
CA439825
LOCUS       CA439825                655 bp    mRNA    linear   EST 08-NOV-2002
DEFINITION  UI-H-DI0-auz-j-14-0-UI.s1 NCI_CGAP_DI0 Homo sapiens cDNA clone
            UI-H-DI0-auz-j-14-0-UI 3', mRNA sequence.
ACCESSION   CA439825
VERSION     CA439825.1  GI:24804245
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 655)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
            sequence: 1-22, >AT_rich#Low_complexity (matched compliment)
            86-110, >POLY_A#Simple_repeat (matched compliment) 111-400, >ALU
            (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .655
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-DI0-auz-j-14-0-UI"
                     /tissue_type="Lung Focal Fibrosis"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_DI0"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_DI0 is a cDNA library containing the following
                     tissue(s): A pool of Lung Focal Fibrosis. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     ATACGCGGTC.
                     TAG_TISSUE=lung with fibrosis
                     TAG_LIB=UI-H-DI0
                     TAG_SEQ=ATACGCGGTC"
ORIGIN
```

```
  Query Match           10.0%;  Score 199.2;  DB 6;  Length 655;
  Best Local Similarity   82.1%;  Pred. No. 4.2e-23;
  Matches  266;  Conservative    0;  Mismatches   53;  Indels    5;  Gaps    3;

Qy       1197 TTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTG 1256
                | ||  |  |||| |    | ||||| ||||| |||||||||| |||||||||  || ||
Db         78 TAGAGCAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGCTCTG 137

Qy       1257 TCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 1316
                ||||||||||||||||| ||||||  ||  ||||||||||||||||||||||||||||||||||||
Db        138 TCACCCAGGCTGGAGTGCAGTGG-TGCAGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 196

Qy       1317 TTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAG--GCGCATGCACC 1372
                ||||||| ||| |||||  |||||| |||||||| | |||  ||||||   |
Db        197 TTCAAGCCATTTTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTGCAGGTGCGCATCACCA 256

Qy       1373 ACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTG 1432
                | || ||||||||||||||||| ||||||||  | ||||| |||||||| ||||||| |
Db        257 CGCCTGGCTAATTTTTGTATTTTGAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGG 316

Qy       1433 CCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTAC 1492
                || ||||||||| | ||| ||||||||||||||||||||||||||||||| || |||||||
Db        317 TCTCGAACTCCTGGCCTCAAGTGATCCACCCACCTCAGCCTCCCAAAGTACAGGGATTAT 376

Qy       1493 AGGCATGAGCCACCGTGCCCAGCC 1516
                |||| || ||||| |||||| |||
Db        377 AGGCGTGCGCCACTGTGCCCGGCC 400


RESULT 13
BZ603010/c
LOCUS       BZ603010                694 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHAAX35TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-6F22, genomic survey sequence.
ACCESSION   BZ603010
VERSION     BZ603010.1  GI:31511472
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 694)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-6F22"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN
```

```
  Query Match             9.9%;  Score 199;  DB 9;  Length 694;
  Best Local Similarity   79.3%;  Pred. No. 4.5e-23;
  Matches  275;  Conservative    0;  Mismatches   65;  Indels    7;  Gaps    3;

Qy       1194 TATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACT 1253
              | |||  || | |||| |    | ||||| ||||| |||||||||| ||||||||| |||
Db        371 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTACT 312

Qy       1254 TTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCC 1313
              ||||||||||||||||  |||||||     |||||||||||||||||||||||||
Db        311 CTGTCACCCAGGCTGGAGGGCAGTGGC--GCATCTCGGCTCACTGCAACCTCTGCCTCCT 254

Qy       1314 AGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCAC 1371
              || |||||||||||||| ||  ||||| |||||| |||||| |||||||||| |  ||
Db        253 GGGCTCAAGCGATTCTCTTGCCTCAGCCTCCCGAGTAGCTGTGATTACAGGCACCCGCCA 194

Qy       1372 C---ACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGG 1428
              |    ||| ||  | |||||||||||||||||||||||| |||||||||||||| |||||
Db        193 CCATGCCAGGCTAATTTTTTGTATTTTTAGTAGAGACAGGGTTTCGCCATGTTGCCCAGG 134

Qy       1429 CTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGA 1488
              || | || |||||||||| |||||   |||  |||| ||||||||||||||  ||||||
Db        133 CTGGTCTGGAACTCCTGACCTCAGGCAATCTGCCCAACTCAGCCTCCCAAAGTGCTGGGA 74

Qy       1489 TTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA 1535
              ||||||| |||||||||| |||||| |    ||  ||  |   |||
Db          73 TTACAGGCCTGAGCCACCGCGCCCAGGCAACATTTCCATTTTTAGTA 27


RESULT 14
BU735183
LOCUS       BU735183                467 bp    mRNA    linear   EST 09-OCT-2002
DEFINITION  UI-E-DW0-agk-o-18-0-UI.s1 UI-E-DW0 Homo sapiens cDNA clone
            UI-E-DW0-agk-o-18-0-UI 3', mRNA sequence.
ACCESSION   BU735183
VERSION     BU735183.1  GI:23663846
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 11-300, >ALU (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-DW0-agk-o-18-0-UI"
                     /tissue_type="lens"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-DW0"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-DW0 is a cDNA library containing the following
                     tissue(s): lens. The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was ligated to an EcoR I adaptor, digested with Not
                     I, and cloned directionally into pT7T3-Pac vector. The
                     oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is CGATTAGCGA. This library
                     was created for the program, Gene Discovery in the Visual
                     System, supported by National Eye Institute (NEI).
                     TAG_TISSUE=human lens
                     TAG_LIB=UI-E-DW0
                     TAG_SEQ=CGATTAGCGA"
ORIGIN

  Query Match             9.9%;  Score 198.8;  DB 5;  Length 467;
  Best Local Similarity   81.2%;  Pred. No. 5.3e-23;
  Matches  268;  Conservative    0;  Mismatches   57;  Indels    5;  Gaps    3;

Qy       1220 TTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGG 1279
              ||| ||||| |||||||||| || ||||||| || |||||||||||||||||| ||||||
Db          1 TTTTTTTTTTTTTTTTTTGAGACGGAGTCTCCCTCTGTCACCCAGGCTGGAGTGCAGTGG 60

Qy       1280 CTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAG 1337
              |      |||||||||||||||||||| ||||||  ||||||||| |||||||||  ||||
Db         61 C-ACCATCTCGGCTCACTGCAACCTCCGCCTCCTGGGTTCAAGCAATTCTCCTGCCTCAG 119

Qy       1338 CTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTATTTT 1395
              | ||| ||  ||||||||||||||||||  ||  |  ||  || ||||||| ||||||||
Db        120 CCTCCTGAGAAGCTGGGATTACAGGCGCCCGCCTCCACACCTGGCTAATTTCTGTATTTT 179

Qy       1396 TAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTG 1455
              ||||| |||||| |||||||||||||| |||||||  | || ||||||  || | |||||||
Db        180 TAGTAAAGACAGGGTTTCGCCATGTTGGCCAGGCTGGTCTGGAACTCTTGTCCTCAGGTG 239

Qy       1456 ATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGC 1515
              |||| |||||||| |||||||||||| ||| |||||||||  |||||||||  |||  ||
Db        240 ATCCGCCCACCTCGGCCTCCCAAAGCGCTGTGATTACAGGCGTGAGCCACCGCGCCTGGC 299

Qy       1516 CTGTTTTCTCAGATCCTGTATTTGTTTCTG 1545
              ||  ||| | || |  | || |   | |||
Db        300 CTTGTTTGTTAGGTTTTTTACTACCTGCTG 329


RESULT 15
AQ469410/c
LOCUS       AQ469410                591 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  CITBI-E1-2596F16.TR CITBI-E1 Homo sapiens genomic clone 2596F16,
            genomic survey sequence.
ACCESSION   AQ469410
VERSION     AQ469410.1  GI:4651288
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2596F16.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
```

```
           Department of Eukaryotic Genomics
           The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850
           Tel: 301 838 0200
           Fax: 301 838 0208
           Email: hbe@tigr.org
           Clones are available from Research Genetics (info@resgen.com). BAC
           end search page:
           http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
           Seq primer: M13 Reverse
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2596F16"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
ORIGIN

  Query Match             9.9%;  Score 198.8;  DB 9;  Length 591;
  Best Local Similarity   85.6%;  Pred. No. 5e-23;
  Matches  256;  Conservative    0;  Mismatches   38;  Indels    5;  Gaps    3;

Qy       1220 TTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGG 1279
              ||| ||||| | |||||||| || |||||||||| |||| |||||||||||||| ||||||
Db        590 TTTTTTTTTTTNTTTTTGAGACGGAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGG 531

Qy       1280 CTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAG 1337
              | | | ||||||||||||||||||||  |||||||||||||| ||||||| ||   |||
Db        530 C-GCGATCTCGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGATTCTCGTGCCTCAA 472

Qy       1338 CTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATTTTTA 1397
              | |||| | |||||||||||||||| ||  || | ||||||||||||||||||||||||
Db        471 CCTCCCTAGTAGCTGGGATTACAGGTGC--GCGCTACCATGCCTAATTTTTGTATTTTTA 414

Qy       1398 GTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGAT 1457
              |||||||||| ||||| |||||||| ||||||| | || |||||||||| |||||||||
Db        413 GTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGAT 354

Qy       1458 CCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCC 1516
              |||||| |||  |||||||||||  || ||||||||||||||||||||  |||  |||
Db        353 CCACCCGCCTTGGCCTCCCAAAGTGCTAGGATTACAGGCATGAGCCACCACGCCTGGCC 295


RESULT 16
AA515728
LOCUS       AA515728                439 bp    mRNA    linear   EST 19-AUG-1997
DEFINITION  ng70f04.s1 NCI_CGAP_Lip2 Homo sapiens cDNA clone IMAGE:940159
            similar to contains Alu repetitive element;contains element PTR5
            repetitive element ;, mRNA sequence.
ACCESSION   AA515728
VERSION     AA515728.1  GI:2255328
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 634   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:940159"
                     /tissue_type="liposarcoma"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Lip2"
                     /note="Vector: pAMP10; mRNA made from liposarcoma, cDNA
                     made by oligo-dT priming. Non- directionally cloned.
                     Size-selected on agarose gel, average insert  size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
ORIGIN

  Query Match             9.9%;  Score 198.6;  DB 1;  Length 439;
  Best Local Similarity   84.1%;  Pred. No. 5.9e-23;
  Matches  260;  Conservative    0;  Mismatches   44;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
              ||||||  ||||| ||||||||||||  |||||||||| |||||||||||||||| | ||
Db         58 TCTTTTCTTTTTTTTTTTTTTTGAAATGGAGTCTCACTCTGTCACCCAGGCTGGAATGCA 117

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
              |||||   | |||| |||||||||||||| ||||||||||||||| ||||||||||
Db        118 GTGGC-ACGATCTCAGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGAGATTCTCCTGCC 176

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAG--GCGCATGCACCACCATGCCTAATTTTTGTA 1391
              |||||  |||||| |||||||||||||  || ||    |   | || ||||||||||||
Db        177 TCAGCCTCCCGAGTAGCTGGGATTACAGCTGCACACCACCAGGCCTGGCTAATTTTTGTA 236

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
              |||||||||||| | | |||||||||||||| ||||||| | ||  |||||||||| |||
Db        237 TTTTTAGTAGAGCCGGGGTTTCGCCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 296

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
              ||||||| |||||||||||||||||||  |||||||| ||||||||||||||    ||
Db        297 AGTGATCCGCCCACCTCAGCCTCCCAAAGTGCTGGGATTATAGGCATGAGCCACCACACC 356

Qy       1512 CAGCCTGTT 1520
              ||||||  ||
Db        357 CAGCCTATT 365


RESULT 17
AA904211
LOCUS       AA904211                455 bp    mRNA    linear   EST 21-APR-1998
DEFINITION  od88e02.s1 NCI_CGAP_Br5 Homo sapiens cDNA clone IMAGE:1377338
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AA904211
VERSION     AA904211.1  GI:3039334
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,
            Ph.D. student, Rodrigo F. Chuaqui, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 579   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1377338"
                     /sex="female"
                     /tissue_type="infiltrating ductal carcinoma"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Br5"
                     /note="Organ: breast; Vector: pAMP10; mRNA made from
                     infiltrating ductal carcinoma, cDNA made by oligo-dT
                     priming.  Non-directionally cloned.  Size-selected on
                     agarose  gel, average insert size 600 bp.  "
ORIGIN

  Query Match             9.9%;  Score 198.4;  DB 1;  Length 455;
  Best Local Similarity   82.2%;  Pred. No. 6.3e-23;
  Matches  264;  Conservative    0;  Mismatches   52;  Indels    5;  Gaps    3;

Qy       1208 TTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCT 1267
              ||| | |    |   | |||| ||||||| |||||||||||||| | ||||||||||||||
Db         25 TTCGCGGCCGCTNCGACTTTTTTTTTTTTTGGAAACAGAGTCTCGCGTTGTCACCCAGGCT 84

Qy       1268 GGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 1327
              ||||| |||||||  |  ||| |||||||||||||||| |||||||||||||||||| |||
Db         85 GGAGTGCAGTGGC-ATAATCTTGGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGCAATT 143

Qy       1328 CTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAA 1383
              ||||||  ||||| ||| || |||||||||||||||||| ||  |  |  ||  || ||||
Db        144 CTCCTGCCTCAGCCTCCAGAGTAGCTGGGATTACAGGCACACACCACCACACCTGGCTAA 203

Qy       1384 TTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCC 1443
              |||||||||||||||||||||| | ||||| | |||||| ||||||| | ||  ||||||
Db        204 TTTTTGTATTTTTAGTAGAGACGGGGTTTCACTATGTTGGCCAGGCTGGTCTCAAACTCC 263

Qy       1444 TGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCC 1503
              |||| |||||||||||||| |||  ||||||||||  ||||||||||||  ||||||
Db        264 TGACCTCAGGTGATCCACCCGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCC 323

Qy       1504 ACCGTGCCCAGCCTGTTTTCT 1524
              |||||||| |||||  ||||
Db        324 ACCGTGCCCGGCCTGCCTTCT 344


RESULT 18
AW970571/c
LOCUS       AW970571                 521 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST382652 MAGE resequences, MAGK Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW970571
VERSION     AW970571.1  GI:8160416
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C.,
            Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 279
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGK"
                     /note="Vector: pBluescriptSKm"
ORIGIN

  Query Match             9.9%;  Score 198.4;  DB 1;  Length 521;
  Best Local Similarity   81.4%;  Pred. No. 6.1e-23;
  Matches  267;  Conservative    0;  Mismatches   56;  Indels    5;  Gaps    3;

Qy       1195 ATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTT 1254
              | ||  || | |||| |      |||||| ||||| ||| |||||| || |||||||| ||
Db        504 ACTTTTTTTTTTTTTTTTTTTCCTTTTTTTTTTTTTTTTGTTTTGAGACGGAGTCTCGCTC 445

Qy       1255 TGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCA 1314
              |||| |||||||||||||| |||||| || | ||||||||||||||||| ||||||||||
Db        444 TGTCGCCCAGGCTGGAGTGCAGTGG-TGCGATCTCGGCTCACTGCAAGCTCTGCCTCCTG 386

Qy       1315 GGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--A 1370
              || ||||| ||||||||||  ||||| |||||| | |||| ||||||||||||  ||
Db        385 GGCTCAAGTGATTCTCCTGCCTCAGCCTCCCGAGTTGCTGAGATTACAGGCGCCCGCCAC 326

Qy       1371 CCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
              |   |  | |||||||||||||||||||||||| | ||||||||||||    ||||||
Db        325 CACGCCCGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCGCCATGTTAGGCAGGCT 266

Qy       1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490
              | ||||||||||||||| |||||||||||||| |||| ||||||||||   ||||||||
Db        265 GGTCTTGAACTCCTGACCTCAGGTGATCCACCCGCCTCGGCCTCCCAAAATGCTGGGATT 206

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTG 1518
              |||||||||||||| |||||| ||| |
Db        205 ACAGGCATGAGCCACTGTGCCCGGCCGG 178


RESULT 19
AQ554309/c
LOCUS       AQ554309                 616 bp    DNA     linear   GSS 28-MAY-1999
DEFINITION  RPCI-11-412I4.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-412I4,
            genomic survey sequence.
ACCESSION   AQ554309
VERSION     AQ554309.1  GI:4913486
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
```

```
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .616
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="GDB:7658019"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-412I4"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /clone_lib="RPCI-11"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             9.9%;  Score 198.4;  DB 9;  Length 616;
  Best Local Similarity   84.9%;  Pred. No. 5.8e-23;
  Matches  258;  Conservative    0;  Mismatches   41;  Indels    5;  Gaps    3;

Qy       1219 TTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTG 1278
              |||| ||||| |||||||||| |||||||||  || || ||||||||||||||| |||||
Db        355 TTTTTTTTTTTTTTTTTTTGACACAGAGTCTTGCTCTGCCACCCAGGCTGGAGTGCAGTG 296

Qy       1279 GCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCC--TGTCA 1336
              ||  || ||||||||||||||||||||| |||||||||||||||||||||||||  |||||
Db        295 GC-ATGATCTCGGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGCGATTCTCCTGTGTCA 237

Qy       1337 GCTTCCCGAATAGCTGGGATTACAGGC--GCATGCACCACCATGCCTAATTTTTGTATTT 1394
              || ||| || ||||||||||||||||  |||     |   || | ||||||||||||||
Db        236 GCCTCCAGAGTAGCTGGGATTACAGGCATGCACCACCACACACGGCTAATTTTTGTATTT 177

Qy       1395 TTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGT 1454
              |||||||| || | ||||| |  ||||| ||||||| | |||||||||||||| ||||||
Db        176 TTAGTAGACACGGGGTTTCACGGTGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGT 117

Qy       1455 GATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAG 1514
              ||| | |||| ||| ||||||||||  |||||||||||||||||||||| |||| ||
Db        116 GATTCTCCCATCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCATGCCAAG 57

Qy       1515 CCTG 1518
              ||||
Db          56 CCTG 53


RESULT 20
BM803650
LOCUS       BM803650                990 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6439541 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5520297
            5', mRNA sequence.
ACCESSION   BM803650
VERSION     BM803650.1  GI:19120473
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 990)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12184  row: b  column: 10
            High quality sequence stop: 632.
FEATURES              Location/Qualifiers
     source           1. .990
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:5520297"
                      /tissue_type="leiomyosarcoma"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="NIH_MGC_71"
                      /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 2.1 kb.  "
ORIGIN

  Query Match             9.9%;  Score 198.4;  DB 3;  Length 990;
  Best Local Similarity   77.5%;  Pred. No. 5.2e-23;
  Matches  279;  Conservative    0;  Mismatches   76;  Indels    5;  Gaps    3;

Qy       1220 TTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGG 1279
              ||| ||||| |||||||||| |||||||||| || |||||||||||||||||| ||||||
Db          27 TTTTTTTTTTTTTTTTTTGAGACAGAGTCTCTCTCTGTCACCCAGGCTGGAGTGCAGTGG 86

Qy       1280 CTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAG 1337
              |||| ||  |||| |||||||||||  ||||||||| ||||| ||||||||||  ||||
Db          87 -TGTGATCGTGGCTTACTGCAACCTCCACCTCCCAGGATCAAGTGATTCTCCTGCCTCAG 145

Qy       1338 CTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTATTTT 1395
              | |||||| |||||||||||||||||||  ||   |   | | ||||||||||| ||||
Db         146 CCTCCCGAGTAGCTGGGATTACAGGCGCCCGCCACCATGCCAGGCTAATTTTTGTGTTTT 205

Qy       1396 TAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTG 1455
              ||||||||||   ||||| |||||||| ||||||| | || |||||||||||| |||||||
Db         206 TAGTAGAGACGAGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTG 265

Qy       1456 ATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGC 1515
              |||||||| |||| |||||| ||||  |||||||||||||||||||||||| |||||||
Db         266 ATCCACCCCCCTCGGCCTCCTAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCAGC 325

Qy       1516 CTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTCATTT 1575
              | | ||    ||  | |    | | | ||  | ||   | |   || | |   |||
Db         326 CAGGGTTAATTAATTTTTAAGAAATCTTTCAAATCATCGAGTTTTGATTTTGTTGATTTT 385


RESULT 21
AQ415030/c
LOCUS       AQ415030                711 bp    DNA     linear   GSS 23-MAR-1999
DEFINITION  RPCI-11-203E3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203E3,
            genomic survey sequence.
ACCESSION   AQ415030
VERSION     AQ415030.1  GI:4473999
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-203E3.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
```

```
           The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850
           Tel: 301 838 0200
           Fax: 301 838 0208
           Email: hbe@tigr.org
           Clones are derived from the human BAC library RPCI-11. For BAC
           library availability, please contact Pieter de Jong
           (pieter@dejong.med.buffalo.edu). Clones may be purchased from
           BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
           Research Genet cs (info@resgen.com). BAC end search page:
           http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
           Seq primer: T7
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .711
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7577666"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-203E3"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             9.9%;  Score 198.2;  DB 9;  Length 711;
  Best Local Similarity   78.0%;  Pred. No. 6e-23;
  Matches  277;  Conservative    0;  Mismatches   73;  Indels    5;  Gaps    3;

Qy       1195 ATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTT 1254
                ||||  || | |||| |    | ||||| ||||| |||||||||| |  ||||||  ||
Db        414 ATTTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTTGCTC 355

Qy       1255 TGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCA 1314
                |||| |||||||||||||| |||||||  || ||||||||||||||| ||| |||||||
Db        354 TGTCGCCCAGGCTGGAGTGCAGTGGC-ATGATCTCGGCTCACTGCAAACTCCGCCTCCCG 296

Qy       1315 GGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACC 1372
                |||||||| |||||||||  ||||| |||||| |||||||||||||||| |||  | |
Db        295 GGTTCAAGCAATTCTCCTGTCTCAGCCTCCCGAGTAGCTGGGATTACAGGTGCACCCTAC 236

Qy       1373 --ACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
                  |    |||||||||| |||||||||||  | ||||| |||||||| |||||||
Db        235 GATGCCCAGCTAATTTTTGTTTTTTTAGTAGAGAAGGGGTTTCTCCATGTTGGCCAGGCT 176

Qy       1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490
                | |||||||||||||| |||||||||| ||  |||||||||||||||  |||| |||
Db        175 GGTCTTGAACTCCTGACCTCAGGTGATCCGCCAGCCTCAGCCTCCCAAAGTGCTGGAATT 116

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTG 1545
                |||||||||||||| |||||| |||| |  | |     | ||     | | | ||
Db        115 ACAGGCATGAGCCACTGTGCCCGGCCTATAGTTTTCATTTCTTATGATATCTTTG 61


RESULT 22
AA169245
LOCUS       AA169245                441 bp    mRNA    linear   EST 09-MAR-1998
DEFINITION  zp19d03.s1 Stratagene fetal retina 937202 Homo sapiens cDNA clone
            IMAGE:609893 3' similar to contains Alu repetitive element;contains
            element MER22 repetitive element ;, mRNA sequence.
ACCESSION   AA169245
VERSION     AA169245.1  GI:1747804
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1024   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 323.
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:4625152"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:609893"
                     /sex="mixed"
                     /lab_host="SOLR (kanamycin resistant)"
                     /clone_lib="Stratagene fetal retina 937202"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. Pooled
                     retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR
                     Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
ORIGIN

  Query Match             9.9%;  Score 198;  DB 1;  Length 441;
  Best Local Similarity   84.1%;  Pred. No. 7.4e-23;
  Matches  259;  Conservative    0;  Mismatches   45;  Indels    4;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
                | ||||| ||||| |||||||||| |||||| |||||| |||| | ||||||||||| ||
Db          2 TTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGCCTCACTCTGTCTCTCAGGCTGGAGTGCA 61

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
                |||||  | | |||||||||||||||||||| ||||||  |||||||||  | |||||||
Db         62 GTGGC-GAGATCTCGGCTCACTGCAACCTCCGCCTCCTGGGTTCAAGCAACTCTCCTGCC 120

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC-CATGCCTAATTTTTGTAT 1392
                |||||  ||| ||  |||||||||||||||||||  |||  |||       |||||||||| |
Db         121 TCAGCCTCCTGAGTAGCTGGGATTACAGGCGCCCGCAACACATCCAGCTAATTTTTGTGT 180

Qy       1393 TTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAG 1452
                |||||||||||||| |||||| |||||||| |||||| | || |||||||| ||||||
Db         181 TTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTAACTTCAA 240

Qy       1453 GTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCC 1512
                || |||||||| ||||||||||||||||  |||||||||||||||||||||||||| ||| |
Db         241 GTCATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCATGCTC 300

Qy       1513 AGCCTGTT 1520
                |||   ||
Db         301 AGCAAATT 308


RESULT 23
AQ788367
LOCUS       AQ788367                751 bp    DNA     linear   GSS 03-AUG-1999
DEFINITION  HS_3135_A2_E11_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3135 Col=22 Row=I, genomic survey
            sequence.
ACCESSION   AQ788367
VERSION     AQ788367.1  GI:5695991
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Hominidae; Homo.
REFERENCE  1  (bases 1 to 751)
  AUTHORS  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
           Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
           Hood,L.
  TITLE    Sequence-tagged connectors: A sequence approach to mapping and
           scanning the human genome
  JOURNAL  Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED  10449764
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
           High Throughput Sequencing Center
           University of Washington
           401 Queen Anne Avenue North, Seattle, WA 98109, USA
           Tel: (206) 616-3618
           Fax: (206) 616-3887
           Email: jwallace@u.washington.edu
           Clones may be purchased from Research Genetics (info@resgen.com).
           BAC end Web Server: http://www.htsc.washington.edu
           Plate: 3135  row: I  column: 22
           Seq primer: M13 Reverse
           Class: BAC ends
           High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3135 Col=22 Row=I"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

  Query Match             9.9%;  Score 198;  DB 9;  Length 751;
  Best Local Similarity   85.1%;  Pred. No. 6.4e-23;
  Matches  257;  Conservative    0;  Mismatches   40;  Indels    5;  Gaps    3;

Qy       1219 TTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTG 1278
              |||  ||||| |||||||||| ||||||||||||| |||  |||||||| ||||||||||
Db        189 TTTCTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTTGCCCAGGCTAGAGTACAGTG 248

Qy       1279 GCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCA 1336
              | || | ||||||||||||||||||||||| |||||||| ||||||||||||||  |||
Db        249 G-TGCGATCTCGGCTCACTGCAACCTCTGCTTCCCAGGTCCAAGCGATTCTCCTGCCTCA 307

Qy       1337 GCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTATTT 1394
              || |||||| ||||||||| || |||| | |||  |  ||  || ||||||| |||||||
Db        308 GCCTCCCGAGTAGCTGGGACTATAGGCACGTGCCACCACACCTGGCTAATTTCTGTATTT 367

Qy       1395 TTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGT 1454
              |||||||||||||| |||| |||||||| ||||||| | || ||||||||||||||||||
Db        368 TTAGTAGAGACAGGGTTTTACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACTTCAGGT 427

Qy       1455 GATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAG 1514
              ||||||||   ||  ||||||||||   |||||||||||||||||||||||||||||| |
Db        428 GATCCACCTGTCTTGGCCTCCCAAAGTGTTGGGATTACAGGCATGAGCCACCGTGCCCGG 487

Qy       1515 CC 1516
              ||
Db        488 CC 489


RESULT 24
BU603903/c
LOCUS       BU603903                322 bp    mRNA    linear   EST 20-SEP-2002
DEFINITION  AGENCOURT_8932582 NIH_MGC_141 Homo sapiens cDNA clone IMAGE:6499491
            5', mRNA sequence.
ACCESSION   BU603903
VERSION     BU603903.1  GI:23255662
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 322)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: NCI
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2683  row: n  column: 04
            High quality sequence stop: 321.
FEATURES             Location/Qualifiers
     source          1. .322
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6499491"
                     /tissue_type="mixed (pool of 40 RNAs)"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_141"
                     /note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc);
                     Site_2: SfiI (ggccgcctcggcc); Double-stranded cDNA was
                     prepared from a pool of 40 cell line polyA+ RNAs (bladder
                     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
                     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
                     kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
                     ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
                     gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
                     used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.2-0.5
                     kb size fraction (other fractions present in NIH_MGC_142).
                     Library created in the laboratory of M. Brownstein (NIMH,
                     NIH). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             9.9%;  Score 197.8;  DB 5;  Length 322;
  Best Local Similarity   84.4%;  Pred. No. 8.6e-23;
  Matches  259;  Conservative    0;  Mismatches   42;  Indels    6;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCAC-TTTGTCACCCAGGCTGGAGTAC 1274
              | ||||| |||| |||||||||  ||||||| |||| |||||||||||||||||||| |
Db         309 TTTTTTTTTTTTTTTTTTTTTTGCGACAGAGTTTCACTTTTGTCACCCAGGCTGGAGTGC 250

Qy       1275 AGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG- 1333
              | |||  ||| |||||||||||||||||||||||||||||||||||||| ||||||||||
Db         249 AATGGGCGTGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGT 190

Qy       1334 -TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG---CACCACCATGCCTAATTTTTG 1389
              |||| ||| || |||||||||||||| || ||     |   |  |  |||||||||
Db         189 TTCAGCCTCCTGAGTAGCTGGGATTACAGGTGCCTGACAACCAGGCCCGGTTAATTTTTG 130

Qy       1390 TATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTT 1449
              |||||||||| |||| | ||||| ||| ||||  |||||| | || ||||||||||| |
Db         129 TATTTTTAGTAAAGACGGGGTTTCACCACGTTGGTCAGGCTGGTCTCGAACTCCTGACCT 70

Qy       1450 CAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTG 1509
              |||||||||||||||||| |||||||||||  |||||||||||||| ||||||| |||
Db          69 CAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGTG 10

Qy       1510 CCCAGCC 1516
              |||||||
Db           9 CCCAGCC 3
```

```
RESULT 25
BU588173/c
LOCUS       BU588173                324 bp    mRNA    linear   EST 20-SEP-2002
DEFINITION  AGENCOURT_8952135 NIH_MGC_141 Homo sapiens cDNA clone IMAGE:6386321
            5', mRNA sequence.
ACCESSION   BU588173
VERSION     BU588173.1  GI:23239452
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: NCI
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2584  row: f  column: 18
            High quality sequence stop: 323.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6386321"
                     /tissue_type="mixed (pool of 40 RNAs)"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_141"
                     /note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc);
                     Site_2: SfiI (ggccgcctcggcc); Double-stranded cDNA was
                     prepared from a pool of 40 cell line polyA+ RNAs (bladder
                     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
                     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
                     kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
                     ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
                     gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
                     used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.2-0.5
                     kb size fraction (other fractions present in NIH_MGC_142).
                     Library created in the laboratory of M. Brownstein (NIMH,
                     NIH). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             9.9%;  Score 197.8;  DB 5;  Length 324;
  Best Local Similarity   80.6%;  Pred. No. 8.6e-23;
  Matches  257;  Conservative    0;  Mismatches   57;  Indels    5;  Gaps    2;

Qy       1203 ATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCC 1262
              | ||||| |    | ||||| ||||| |||||   || | ||  |     ||||||||||
Db        321 AGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGAGACAGAGTTTTCACTTTTGTCACCC 262

Qy       1263 AGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAG 1322
              |||||||||| || |||  ||| ||||||||||||||||||||||||||||||||||||
Db        261 AGGCTGGAGTGCAATGGGCGTGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAG 202

Qy       1323 CGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG---CACCACCAT 1377
              ||||||||||  ||||| ||| || ||||||||||||||| || ||     |   |
Db        201 TGATTCTCCTGTTTCAGCCTCCTGAGTAGCTGGGATTACAGGTGCCTGACAACCAGGCCC 142

Qy       1378 GCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTG 1437
              | ||||||||||||||||||||| |||| | ||||| ||| |||| |||||| | || |
Db        141 GGTTAATTTTTGTATTTTTAGTAAAGACGGGGTTTCACCACGTTGGTCAGGCTGGTCTCG 82

Qy       1438 AACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCA 1497
              |||||||||| |||||||||||||||||||| |||||||||||  ||||||||||||||
Db          81 AACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCG 22

Qy       1498 TGAGCCACCGTGCCCAGCC 1516
              |||||||| ||||||||||
Db          21 TGAGCCACTGTGCCCAGCC 3


RESULT 26
AQ528478/c
LOCUS       AQ528478                757 bp    DNA     linear   GSS 18-MAY-1999
DEFINITION  RPCI-11-349N11.TV RPCI-11 Homo sapiens genomic clone
            RPCI-11-349N11, genomic survey sequence.
ACCESSION   AQ528478
VERSION     AQ528478.1  GI:4840591
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 757)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-349N11.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .757
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7633954"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-349N11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             9.9%;  Score 197.8;  DB 9;  Length 757;
  Best Local Similarity   79.1%;  Pred. No. 6.9e-23;
  Matches  273;  Conservative    0;  Mismatches   67;  Indels    5;  Gaps    3;

Qy       1197 TTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTG 1256
              ||||||  | | | |    || || || ||||| |||||||||| |||||||||||| ||
Db         589 TTGATTCAGGTGTTTACCATTTTCTTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTG 530

Qy       1257 TCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 1316
              | |||||||||||||| |||||| || |  || ||||||||||||||| |||| || ||
Db         529 TTACCCAGGCTGGAGTGCAGTGG-TGCGACCTTGGCTCACTGCAACCTCCGCCTACCGGG 471
```

```
Qy       1317 TTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACC 1372
Db        470 TTCAAGCAATTCTCCTGCCTCGGCCTCCCAAGTAGGTGGGATTACAGGTGCCTGCCACCA 411

Qy       1373 ACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTG 1432
Db        410 TGCCCAGCTAATTTTTGTATTTTTAGTAAAGACAGGGTTTCACCATGTTGGCCAGGCTGG 351

Qy       1433 CCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTAC 1492
Db        350 TCTTGAACTCCTGACCTCAAGTGATCCACCCACCTCAGCCTCCCAAAGTGCTGAGATTAC 291

Qy       1493 AGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATT 1537
Db        290 AGGCATGAGCCACTGCACCAGGCCATGTTTACTACTTTCTTCCTT 246
```

RESULT 27
AQ738849/c

LOCUS	AQ738849 896 bp DNA linear GSS 16-JUL-1999
DEFINITION	HS_5386_B1_G06_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=962 Col=11 Row=N, genomic survey sequence.
ACCESSION	AQ738849
VERSION	AQ738849.1 GI:5516371
KEYWORDS	GSS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 896)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED	10449764
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 962 row: N column: 11 Seq primer: T7 Class: BAC ends High quality sequence stop: 896.

FEATURES	Location/Qualifiers
source	1. .896 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="Plate=962 Col=11 Row=N" /sex="male" /clone_lib="RPCI-11 Human Male BAC Library" /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

ORIGIN

Query Match 9.9%; Score 197.8; DB 9; Length 896;
Best Local Similarity 79.1%; Pred. No. 6.7e-23;

Matches 273; Conservative 0; Mismatches 67; Indels 5; Gaps 3;

```
Qy       1201 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCAC 1260
Db        744 TTCTTTTTCTCTAATTTATTTCATTTTTTTTTTTTTTGAGACAGAGTCTCATTCTGTCAC 685

Qy       1261 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 1320
Db        684 CCAGGCTGGAGGGCCGTGGC-GGAATCTCAGTTCACTGCAACCTCTGCCTCTGGGATTCA 626

Qy       1321 AGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGG--CGCATGCACCACCA 1376
Db        625 AGCAATTCTTCTGCCTCGGCCTCCTGAGCAGCTGGGATTATAGGTACGCACCACCACACC 566

Qy       1377 TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTT 1436
Db        565 CGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGACCAGGCTGGTCTT 506

Qy       1437 GAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGC 1496
Db        505 GAACTCCTGACCTCAGGTGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGC 446

Qy       1497 ATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
Db        445 GTGAGCCACCGCGCCCAGCCTCTTTATTTACTTCTTTAATAAATT 401
```

RESULT 28
BX486751

LOCUS	BX486751 493 bp mRNA linear EST 04-SEP-2003
DEFINITION	DKFZp686E18253_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFZp686E18253 5', mRNA sequence.
ACCESSION	BX486751
VERSION	BX486751.1 GI:31950724
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 493)
AUTHORS	Bloecker,H., Boecher,M., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
TITLE	EST (Bloecker,H., Boecher,M., Mewes,H.W., Weil,B., Amid,C., et al.)
JOURNAL	Unpublished (2003)
COMMENT	Contact: MIPS MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de; sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project. No s1 sequence available. This clone (DKFZp686E18253) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES	Location/Qualifiers
source	1. .493 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="DKFZp686E18253" /dev_stage="adult" /lab_host="DH10B" /clone_lib="686 (synonym: hlcc3)" /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"

ORIGIN

Query Match 9.9%; Score 197.6; DB 5; Length 493;
Best Local Similarity 72.3%; Pred. No. 8.3e-23;

```
  Matches  298;  Conservative    0;  Mismatches  109;  Indels    5;  Gaps    3;

Qy       1197 TTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCACTTTG 1256
              ||  |  | |||  |   || ||||| ||||| |||||||||| | || |||||||| ||
Db         28 TTTTTGGTTTTTGTTTTTATTTTTTTTTTTTTTTTTTTTTTGAGATAGGGTCTCACTCTG 87

Qy       1257 TCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 1316
              || |||||||||||| ||||| |  || ||| |||||||||||||||||||||||||| ||
Db         88 TCGCCCAGGCTGGAGTGCAGTGAC-ATGATCTTGGCTCACTGCAACCTCTGCCTCCCTGG 146

Qy       1317 TTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC 1374
              |||||||||||| ||||  ||||| ||| || ||| | || |||||||| || ||  |
Db        147 TTCAAGCGATTCCCCTGCCTCAGCCTCCTGAGTAGTTTGGGTTACAGGCACAAGCCACCA 206

Qy       1375 CATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTG 1432
              |   |  |||||||||| |||||||||||||  | ||||| |||||||| ||||||| |
Db        207 CGCCCAGCTAATTTTTGCATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGG 266

Qy       1433 CCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTAC 1492
              || |||| |||||| |||||||||  ||| |||  |||||||| ||  || |||||||
Db        267 TCTCGAACCCCTGACCTCAGGTGATCTGCCCGCCTTGGCCTCCCACAGTGCTTGGATTAC 326

Qy       1493 AGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTT 1552
              |||||||||||||| |||||||||  |  | |   || |||||||| |||  |
Db        327 AGGCATGAGCCACCATGCCCAGCCGATAATTTGCCATTCTGTATTTATTTGCGTGTTAGA 386

Qy       1553 CATTTCTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTT 1604
              |    |      | |  || |||  || ||| |  | |   || ||   ||
Db        387 GACCAGTTAAAGTTAAGACACTTTCTAAATAGAAGTCTTGTATATGGGGCTT 438


RESULT 29
BZ603705/c
LOCUS       BZ603705                607 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHADP21TR Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-22D18, genomic survey sequence.
ACCESSION   BZ603705
VERSION     BZ603705.1  GI:31512167
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-22D18"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN

  Query Match             9.9%;  Score 197.6;  DB 9;  Length 607;
  Best Local Similarity   84.1%;  Pred. No. 7.9e-23;
  Matches  259;  Conservative    0;  Mismatches   44;  Indels    5;  Gaps    3;

Qy       1219 TTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTG 1278
              |||  || ||  ||| |||    | | | ||||||| |||| ||||||||||||||||||
Db        449 TTTATTTATTTGTTTGTTTACGAAACACTCTCACTCTGTCGCCCAGGCTGGAGTACAGTG 390

Qy       1279 GCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCA 1336
              | |||| ||||||||||||||||||||||||||||||||| |||| |||||||||  |||
Db        389 G-TGTGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTGAAGCAATTCTCCTGCCTCA 331

Qy       1337 GCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTATTT 1394
              || |||||| ||||||||||||||||||||  |    |    | || ||||||||||||||
Db        330 GCCTCCCGAGTAGCTGGGATTACAGGCGCACACCACCATGCCTGGCTAATTTTTGTATTT 271

Qy       1395 TTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGT 1454
              ||||| ||||||| ||||| |||||||| ||||||| | |||||||||||||| ||||||
Db        270 TTAGTGGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGT 211

Qy       1455 GATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAG 1514
              ||||  ||||||||||||||||||  ||||||||||||  ||||||||||||||||||
Db        210 GATCTGCCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGTGCCCAG 151

Qy       1515 CCTGTTTT 1522
              | |   ||
Db        150 CTTCAATT 143


RESULT 30
BU617273
LOCUS       BU617273                730 bp    mRNA    linear   EST 23-SEP-2002
DEFINITION  UI-H-DF0-bep-o-19-0-UI.s1 NCI_CGAP_DF0 Homo sapiens cDNA clone
            UI-H-DF0-bep-o-19-0-UI 3', mRNA sequence.
ACCESSION   BU617273
VERSION     BU617273.1  GI:23283488
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 730)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
            sequence: 1-36, >POLY_A#Simple_repeat (matched compliment) 37-326,
            >ALU (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .730
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-DF0-bep-o-19-0-UI"
                     /tissue_type="Subchondral Bone"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_DF0"
```

/note="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI_CGAP_DF0 is a cDNA library containing the following tissue(s): Subchondral Bone. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTTAAGCGTC.
TAG_TISSUE=subchondral bone
TAG_LIB=UI-H-DF0
TAG_SEQ=GTTAAGCGTC"

ORIGIN

Query Match 9.9%; Score 197.6; DB 5; Length 730;
Best Local Similarity 74.5%; Pred. No. 7.6e-23;
Matches 289; Conservative 0; Mismatches 94; Indels 5; Gaps 3;

```
Qy       1194 TATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACT 1253
             | |||  || | |||| |     ||| ||  ||||| |||| ||| | |||| ||||||
Db          1 TTTTTTTTTTTTTTTTTGGTTTCTATTACTTTTTTTTTTGTTTTAGACAGGGTCTCATT 60

Qy       1254 TTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCC 1313
              |||||||||| ||||||| || ||| |||| |||| ||||||||||||||||||||||||
Db         61 CTGTCACCCAGTCTGGAGTGCAATGG-TGTGATCTCAGCTCACTGCAACCTCTGCCTCCC 119

Qy       1314 AGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGC--GCATGC 1369
               ||||||| ||||| ||||  ||||| |||| | ||||| ||||||||||  | |
Db        120 GAGTTCAAGTGATTCCCCTGCCTCAGCCTCCCAAGTAGCTAGGATTACAGGCATGTACCA 179

Qy       1370 ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGC 1429
              || || || |||||||||||||||||||||||||||  |||| |||||||| ||||||
Db        180 CCCTCCCTGACTAATTTTTGTATTTTTAGTAGAGACAGCATTTCACCATGTTGGCCAGGC 239

Qy       1430 TTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGAT 1489
              | | ||  |||||||| | || |||||||||||| ||||||||||||| |  |||| ||
Db        240 TGGTCTCAAACTCCTGGCCTCGGGTGATCCACCCGCCTCAGCCTCCCAATGTGCTGGAAT 299

Qy       1490 TACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGC 1549
              |||||   |||||||| |  ||  ||||| |||||    |   | ||    | | | | |
Db        300 TACAGTTGTGAGCCACTGCACCTGGCCTGGTTTCTATTACTTTTTAAACATCTTTAATGA 359

Qy       1550 CTTCATTTCTATCTTCTTATTCATTTTG 1577
              || |  | ||| ||| || |    |||
Db        360 TTTAAAATGTATTTTCATAATAGCCTTG 387
```

RESULT 31
CA428248
LOCUS CA428248 741 bp mRNA linear EST 07-NOV-2002
DEFINITION UI-H-DF0-ber-k-13-0-UI.s1 NCI_CGAP_DF0 Homo sapiens cDNA clone UI-H-DF0-ber-k-13-0-UI 3', mRNA sequence.
ACCESSION CA428248
VERSION CA428248.1 GI:24790974
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 741)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA sequence: 1-36, >POLY_A#Simple_repeat (matched compliment) 37-326, >ALU (matched compliment)
Seq primer: M13 FORWARD
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .741
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-DF0-ber-k-13-0-UI"
/tissue_type="Subchondral Bone"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_DF0"
/note="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI_CGAP_DF0 is a cDNA library containing the following tissue(s): Subchondral Bone. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTTAAGCGTC.
TAG_TISSUE=subchondral bone
TAG_LIB=UI-H-DF0
TAG_SEQ=GTTAAGCGTC"

ORIGIN

Query Match 9.9%; Score 197.6; DB 6; Length 741;
Best Local Similarity 74.5%; Pred. No. 7.5e-23;
Matches 289; Conservative 0; Mismatches 94; Indels 5; Gaps 3;

```
Qy       1194 TATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACT 1253
             | |||  || | |||| |     ||| ||  ||||| |||| ||| | |||| ||||||
Db          1 TTTTTTTTTTTTTTTTTGGTTTCTATTACTTTTTTTTTTGTTTTAGACAGGGTCTCATT 60

Qy       1254 TTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCC 1313
              |||||||||| ||||||| || ||| |||| |||| ||||||||||||||||||||||||
Db         61 CTGTCACCCAGTCTGGAGTGCAATGG-TGTGATCTCAGCTCACTGCAACCTCTGCCTCCC 119

Qy       1314 AGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGC--GCATGC 1369
               ||||||| ||||| ||||  ||||| |||| | ||||| ||||||||||  | |
Db        120 GAGTTCAAGTGATTCCCCTGCCTCAGCCTCCCAAGTAGCTAGGATTACAGGCATGTACCA 179

Qy       1370 ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGC 1429
              || || || |||||||||||||||||||||||||||  |||| |||||||| ||||||
Db        180 CCCTCCCTGACTAATTTTTGTATTTTTAGTAGAGACAGCATTTCACCATGTTGGCCAGGC 239

Qy       1430 TTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGAT 1489
              | | ||  |||||||| | || |||||||||||| ||||||||||||| |  |||| ||
Db        240 TGGTCTCAAACTCCTGGCCTCGGGTGATCCACCCGCCTCAGCCTCCCAATGTGCTGGAAT 299

Qy       1490 TACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGC 1549
              |||||   |||||||| |  ||  ||||| |||||    |   | ||    | | | | |
Db        300 TACAGTTGTGAGCCACTGCACCTGGCCTGGTTTCTATTACTTTTTAAACATCTTTAATGA 359

Qy       1550 CTTCATTTCTATCTTCTTATTCATTTTG 1577
              || |  | ||| ||| || |    |||
Db        360 TTTAAAATGTATTTTCATAATAGCCTTG 387
```

RESULT 32
AQ415537/c
LOCUS AQ415537 690 bp DNA linear GSS 23-MAR-1999

```
DEFINITION  RPCI-11-177C16.TV RPCI-11 Homo sapiens genomic clone
            RPCI-11-177C16, genomic survey sequence.
ACCESSION   AQ415537
VERSION     AQ415537.1  GI:4474506
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-177C16.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7567647"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-177C16"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             9.9%;  Score 197.4;  DB 9;  Length 690;
  Best Local Similarity   78.3%;  Pred. No. 8.3e-23;
  Matches  275;  Conservative    0;  Mismatches   71;  Indels    5;  Gaps    3;

Qy       1179 AATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTG 1238
Db        379 AGTGGTTAGAACAGGTCTCACAGAAGTGGTATTGCCTTTTTTTGTTCTTTTTTTTGTTTG 320

Qy       1239 AAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTG 1298
Db        319 AGACAGAGTTTCACTCTGTCGCCCAGGCTGGAGTGCAGTGG-TACGATCTTGGCTCACTG 261

Qy       1299 CAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGAT 1356
Db        260 CAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGAT 201

Qy       1357 TACAGGCGCATGC--ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCG 1414
Db        200 TACAGGTGCACACCACCACTCCCGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTTG 141

Qy       1415 CCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTC 1474
Db        140 CCATGTTGTCCAGGCTGGTCTTGAAATCCTGACCTCAGCTGATCCACCTTCCTCAGCCTC 81

Qy       1475 CCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTC 1525
Db          80 CCAAAGTGCTGGGATTACAGGCATAAGCCACCGCGCCTGGCTGATATTGTC 30
```

```
RESULT 33
CL423123/c
LOCUS       CL423123                720 bp    DNA     linear   GSS 12-MAR-2004
DEFINITION  RP11-429A8, T7 RPCI-11 Human Male BAC Library Homo sapiens genomic
            clone RP11-429A8, genomic survey sequence.
ACCESSION   CL423123
VERSION     CL423123.1  GI:45388254
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Zhao,S. and Yonescu,R.
  TITLE     Direct Submission of BAC End Sequence of CCAP clones
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Yonescu, R
            NCI/NIH
            9000 Rockville Pike, Bethesda, MD 20892, USA
            Tel: 301-402-2008
            Fax: 301-402-1204
            Email: yonescur@pop.nci.nih.gov
            Plate: 429  row: A  column: 8
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-429A8"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             9.9%;  Score 197.4;  DB 10;  Length 720;
  Best Local Similarity   75.7%;  Pred. No. 8.2e-23;
  Matches  284;  Conservative    0;  Mismatches   86;  Indels    5;  Gaps    3;

Qy       1147 GTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGT 1206
Db        656 GTGAGAGGGGACGGGCTGCAGCGCAGAGGTTGGTTGGGTGGCTGCAGGAGCACATGGAGT 597

Qy       1207 TTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGC 1266
Db        596 GTCTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTTCTGTGTCACCCAGGC 537

Qy       1267 TGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGAT 1326
Db        536 TGGAGTGCAGTGG-TGCAATCTTGGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGAT 478

Qy       1327 TCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTA 1382
Db        477 TCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGTCATGTGCCAAAACGCCCGGCTA 418

Qy       1383 ATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTC 1442
Db        417 ATTTTTGTATTTTTAGTACAGACAGAGTTTCACCATGTTGGCCAAGCTGGTCTTGAATTC 358

Qy       1443 CTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGC 1502
Db        357 CTGACCTCAAGTGATCCGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAAACGTGAGC 298

Qy       1503 CACCGTGCCCAGCCT 1517
```

```
Db        297 CACCATGCCCGGCCT 283


RESULT 34
AQ471471/c
LOCUS       AQ471471                660 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  CITBI-E1-2600N15.TF CITBI-E1 Homo sapiens genomic clone 2600N15,
            genomic survey sequence.
ACCESSION   AQ471471
VERSION     AQ471471.1  GI:4655125
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2600N15"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
ORIGIN

  Query Match             9.9%;  Score 197.2;  DB 9;  Length 660;
  Best Local Similarity   77.6%;  Pred. No. 9e-23;
  Matches  264;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    2;

Qy       1219 TTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTG 1278
             |||| |||||  |||||||| |||| |||||||| |||||||||||||||||| |||||
Db        583 TTTTCTTTTTTATTTTTTTGAGACAGGGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTG 524

Qy       1279 GCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCA 1336
             || ||| |||||||||||||||| |||||||||   |||||||  |||||| |||  |||
Db        523 GC-GTGATCTCGGCTCACTGCAATCTCTGCCTCATGGGTTCAAATGATTCTTCTGCCTCA 465

Qy       1337 GCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATTTTT 1396
             |  |||||| |||||| || |||||| |      |    |  | ||||||||||||||||
Db        464 GTCTCCCGAGTAGCTGAGACTACAGGTGGGCCATCACGCTGGGCTAATTTTTGTATTTTT 405

Qy       1397 AGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGA 1456
              |||||||||| ||||| ||| |||| |||||||| |||| ||||||||||| ||||||||
Db        404 GGTAGAGACAGGGTTTCACCACGTTGGCCAGGCTGGCCTCGAACTCCTGACCTCAGGTGA 345

Qy       1457 TCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCC 1516
             ||||||||||   |||||||||  |||||||||||||  ||||||||| ||||| ||
Db        344 TCCACCCACCTTGACCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCCGGCA 285

Qy       1517 TGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATT 1556
                || |   ||  ||  ||     || || |  | |||


Db        284 GAAATTTTTTCATTTTGGTTTAAAGACTAAAACTGTAATT 245


RESULT 35
AA832175
LOCUS       AA832175                410 bp    mRNA    linear   EST 18-MAR-1998
DEFINITION  oc91c04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1357062 3'
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AA832175
VERSION     AA832175.1  GI:2905274
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1357   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 402.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1357062"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

  Query Match             9.8%;  Score 196.8;  DB 1;  Length 410;
  Best Local Similarity   77.6%;  Pred. No. 1.2e-22;
  Matches  264;  Conservative    0;  Mismatches   72;  Indels    4;  Gaps    2;

Qy       1220 TTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGG 1279
             ||| ||||| |||||||||| || ||||||  || |||  |||||||||||| ||||||
Db          1 TTTTTTTTTTTTTTTTTTGAGACGGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTGG 60

Qy       1280 CTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAG 1337
             |    |||||||||||||||||||| |||||||||||||||| |||||||||  ||||
Db         61 CACGAGTCTCGGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGTGATTCTCCTGCCTCAG 120

Qy       1338 CTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTATTTT 1395
             | |||| | |||||||||||||| || |||||  || |   |  ||| ||||||||||||
```

```
Db        121 CCTCCCTAGTAGCTGGGATTACAAGCACATGCCACAACGCCCAGCTAGTTTTTGTATTTT 180

Qy       1396 TAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTG 1455
              |||||||||| | ||||||| |||||| ||||||| | || |||||||| | |||||||
Db        181 TAGTAGAGACCGGGTTTCGCTATGTTGGCCAGGCTGGTCTCGAACTCCTGTCCTCAGGTG 240

Qy       1456 ATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGC 1515
              ||||||||| |||| | |||||||||  ||||||||||||  |||||||| ||||   ||
Db        241 ATCCACCCGCCTCGGACTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGTGCTTGGC 300

Qy       1516 CTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCAT 1555
              | |  |    |     | |  | | |   |||| |  | |
Db        301 CCGAATCAGGAATAATTCTGATGGCTAAGGAAGACAGCTT 340
```

RESULT 36
CR861212

```
LOCUS       CR861212               2978 bp    mRNA    linear   HTC 12-NOV-2004
DEFINITION  Pongo pygmaeus mRNA; cDNA DKFZp459D133 (from clone DKFZp459D133).
ACCESSION   CR861212
VERSION     CR861212.1  GI:55733198
KEYWORDS    HTC.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pongo.
REFERENCE   1  (bases 1 to 2978)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Agowa (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp459D133) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp459D133
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .2978
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459D133"
                     /tissue_type="cortex"
                     /clone_lib="459 (synonym: pcor1). Vector pSport1_Sfi; host
                     DH10B; sites SfilA + SfilB"
                     /dev_stage="adult"
                     /note="tuftelin (Homo sapiens), differentially spliced"
     gene            1. .2978
                     /gene="DKFZp459D133"
     CDS             join(63. .311,311. .1069)
                     /gene="DKFZp459D133"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH93283.1"
                     /db_xref="GI:55733199"
                     /translation="MNGTRNWCTLVDVHPEDQAAAGRKTYAMVSGHSAGHSLASELVE
                     SHDGHEEIIKVYLKGRSGEKMIHEKNINQLKSEVQYIQEVVLEKPNGFSQSPTALYSS
                     LPEVDTCINEDVESLRKTVQDLLAKLQEAERQHQSDRVAFEVTLSRYQREAEERNVSL
                     QREEDRVEQKEAEVGELQRRLLGMETEHQALLAKVREGEVALEELRSNNADCQAEQEK
                     AATLEKEVAGLREKIHHLDDMLKSQQRKVRQMIEQLQNSKAVIQSKDTTIQELKEKIA
                     YLEAENLEMHDRMEHLIEKQISHGNFSTQARAKTENPGSIRISKPPSPKPKPVIRVVE
                     T"
ORIGIN
```

Query Match 9.8%; Score 196.8; DB 4; Length 2978;
Best Local Similarity 70.7%; Pred. No. 7.2e-23;
Matches 290; Conservative 0; Mismatches 117; Indels 3; Gaps 2;

```
Qy       1202 TATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACC 1261
              | | || | .  | | ||||| ||||| |||||||||| |||||||||| || |||||||
Db       2298 TTTTTTGTTGTTGTTGTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCTCTCTGTCACC 2357

Qy       1262 CAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAA 1321
              |||||||||| ||||||| ||| |||| |||| ||||||||||||||||  | || ||
Db       2358 CAGGCTGGAGTGCAGTGGC-GTGATCTCAGCTCCCTGCAACCTCTGCCTCCTGGATTTAA 2416

Qy       1322 GCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 1379
              |||||||| |||  ||||| ||| || ||||||||| |||||| |       |   | ||
Db       2417 GCGATTCTTCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGTGTGCCACCACGCCTGG 2476

Qy       1380 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 1439
              || ||||| |||||||||||||||||  | ||||| ||||  || | ||||| | || |||
Db       2477 CTGATTTTAGTATTTTTAGTAGAGATGGGGTTTCACCATACTGGCTAGGCTGGTCTCGAA 2536

Qy       1440 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 1499
              |||||||| |||||||||||||||||||| || ||||||||  |||||||||||||| ||
Db       2537 CTCCTGACCTCAGGTGATCCACCCACCTCGGCTTCCCAAAGTGCTGGGATTACAGGCTTG 2596

Qy       1500 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCT 1559
              |||||| | |||  |||  |  | |        |    || || ||  |||||| |    |
Db       2597 AGCCACTGCGCCTGGCCCATGGTGTTTTTCTTTAGGGTTCTTCCTACAGCCTTGAGGAGT 2656

Qy       1560 ATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAA 1609
              | |    ||     | |       ||    | || ||  ||   |||||
Db       2657 AGATAGGCATCAGAGTATGGTACTATAGGAATCAGAAAAATTCAAAACAA 2706
```

RESULT 37
CB154854/c

```
LOCUS       CB154854               634 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION  K-EST0212920 B2N807043 Homo sapiens cDNA clone B2N807043-26-E12 5',
            mRNA sequence.
ACCESSION   CB154854
VERSION     CB154854.1  GI:28139955
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 634)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 26  row: E  column: 12
            High quality sequence stop: 634.
FEATURES             Location/Qualifiers
     source          1. .634
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="B2N807043-26-E12"
                     /sex="M"
                     /lab_host="Top10F'"
                     /clone_lib="B2N807043"
                     /note="Organ: Brain; Vector: pCNS-D2; Site_1: EcoRI;
                     Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
```

intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dT-selected mRNA by priming with dT-tailed vector. The dT-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

ORIGIN

Query Match 9.8%; Score 196.6; DB 6; Length 634;
Best Local Similarity 79.8%; Pred. No. 1.1e-22;
Matches 269; Conservative 0; Mismatches 64; Indels 4; Gaps 3;

```
Qy       1210 CTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGG 1269
             ||   | | |||| ||||| |||||||||| |  |||||||||| |||| ||||||||||
Db        353 CTGTAACCATTTTCTTTTTCTTTTTTTTGAGATGGAGTCTCACTCTGTCGCCCAGGCTGG 294

Qy       1270 AGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCT 1329
             ||| ||||| |     |||||||||||||||||||||||||||  |||||||||||||| |
Db        293 AGTGCAGTGAC-ACAATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTGT 235

Qy       1330 CCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCA-CCATGCCTAATTT 1386
             ||||  ||||| ||| || ||||||||||| ||||   ||| |||  |  | |||||||
Db        234 CCTGCCTCAGCCTCCTGAGTAGCTGGGATTATAGGCATGTGCCCCATGCCCGGCTAATTT 175

Qy       1387 TTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGA 1446
             |||||||||||||||||| | ||||| |||| |||| |||| | | || ||||||||||
Db        174 TTGTATTTTTAGTAGAGACGGGGTTTCACCATATTGATCAGGTTGGTCTCGAACTCCTGA 115

Qy       1447 CTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACC 1506
             | |||||||||||||||| ||||||||||||||  |||||||||||||| |||||||||
Db        114 CCTCAGGTGATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACC 55

Qy       1507 GTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTC 1543
               ||| ||||||   |   ||  | |  |  ||| | |
Db         54 ATGCTCAGCCTTGGTAGCCATTTTCATTGATTGATGC 18
```

RESULT 38
AQ506289
LOCUS AQ506289 732 bp DNA linear GSS 29-APR-1999
DEFINITION RPCI-11-299L10.TV RPCI-11 Homo sapiens genomic clone RPCI-11-299L10, genomic survey sequence.
ACCESSION AQ506289
VERSION AQ506289.1 GI:4711036
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 732)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-299L10.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .732
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7614705"
/db_xref="taxon:9606"
/clone="RPCI-11-299L10"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"

ORIGIN

Query Match 9.8%; Score 196.6; DB 9; Length 732;
Best Local Similarity 75.0%; Pred. No. 1.1e-22;
Matches 285; Conservative 0; Mismatches 90; Indels 5; Gaps 3;

```
Qy       1201 TTATGTTTTCTCAGATCTTTTTATTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCAC 1260
             || | |    |  |    ||||| ||||| |||| ||||||| |  || || |||
Db        237 TTCTTTAAGGTTTGNCTTTTTTTTTTTTTTTTTTCTTGAGACAGAGTATTGCTCTGACAC 296

Qy       1261 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 1320
             |||||||  |||| |||||| |||| ||| ||||||||||||||  | |||  | ||||
Db        297 CCAGGCTAGAGTGCAGTGG-TGTGATCTTGGCTCACTGCAACCTCCACTTCCTGGATTCA 355

Qy       1321 AGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA-- 1376
             ||| |||||||||  |  || |||||| |||||||| ||||||| |||||  |  ||
Db        356 AGCAATTCTCCTGCCTTGGCCTCCCGATTAGCTGGGGTTACAGGCACATGCCACCACACC 415

Qy       1377 TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTT 1436
             || |||||||||||||| |||||||||  | ||||| |||||||| ||||||| | |||
Db        416 TGGCTAATTTTTGTATTCTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTT 475

Qy       1437 GAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGC 1496
             |||||||||| |||||||||||||||||||||||| ||||||||  |||| |||||||||
Db        476 GAACTCCTGACCTCAGGTGATCCACCCACCTCAGCTTCCCAAAGTGCTGGAATTACAGGC 535

Qy       1497 ATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATT 1556
             ||||||||| |  ||| ||| | ||| |||  |    ||    ||||  |      | ||
Db        536 ATGAGCCACTGCACCCGGCCAGGTTTGTCAATTTTGATAACCTTTTCAAAGAAACACCTT 595

Qy       1557 TCTATCTTCTTATTCATTTT 1576
             |   | ||  ||  |  ||||
Db        596 TTGGTTTTGTTGATTTTTTT 615
```

RESULT 39
AG182560/c
LOCUS AG182560 669 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-055M11.TJ, genomic survey sequence.
ACCESSION AG182560
VERSION AG182560.1 GI:16712240
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 669)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-055M11.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match             9.8%;  Score 196.4;  DB 10;  Length 669;
  Best Local Similarity   81.5%;  Pred. No. 1.2e-22;
  Matches  277;  Conservative    0;  Mismatches   56;  Indels    7;  Gaps    4;

Qy       1194 TATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACT 1253
              ||||   |    |||  |     | | ||| ||||| |||||||||| ||||||||| ||
Db        457 TATTCCCTGTAATTTCTTTTCTTTTCTTTCTTTTTTTTTTTTTTTGAGACAGAGTCTCGCT 398

Qy       1254 TTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCC 1313
               |||||||||||||| ||| |||||||  || |||||||||| ||||||||| |||||
Db        397 CTGTCACCCAGGCTGCAGTGCAGTGGC-ATGATCTCGGCTCATTGCAACCTCCGCCTCCA 339

Qy       1314 AGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC-- 1369
               ||||||||||||||||||  ||||| ||| || ||||||||||||||| |||  |
Db        338 GGGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGTGCACACCA 279

Qy       1370 ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGC 1429
               |   | || |||||||||||||||||||||||||||||| ||||| |||||||| ||| ||
Db        278 CCACGCCTGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAAGC 219

Qy       1430 TTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGAT 1489
               | | ||||||||||||  | ||  |||||||||||  |||||||||||||||| |||||||
Db        218 TGGTCTTGAACTCCTGGCCTC--GTGATCCACCCGCCTCAGCCTCCCAAAGTACTGGGAT 161

Qy       1490 TACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGAT 1529
               |||||||||||||||||||||||| |||| ||  ||  |||
Db        160 TACAGGCATGAGCCACCGTGCCCGGCCTATTCCCTGTGAT 121


RESULT 40
BZ612069/c
LOCUS       BZ612069                751 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHADP28TR Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-22F8, genomic survey sequence.
ACCESSION   BZ612069
VERSION     BZ612069.1  GI:31520630
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-22F8"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN

  Query Match             9.8%;  Score 196.4;  DB 9;  Length 751;
  Best Local Similarity   80.2%;  Pred. No. 1.2e-22;
  Matches  268;  Conservative    0;  Mismatches   61;  Indels    5;  Gaps    3;

Qy       1197 TTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTG 1256
              ||  ||  |||||      | ||| | ||||| |||||||||| || ||| ||  || ||
Db        436 TTTTTTGGGTTTTTGTTTGTTTTTGTTTTTTTTTTTTTTTTGAGACGGAGCCTTGCTCTG 377

Qy       1257 TCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 1316
              ||||||||||||||||| ||||||  |||  ||| |||||||||||  ||||||||||||  ||
Db        376 TCACCCAGGCTGGAGTGCAGTGG-TGTAATCTTGGCTCACTGCAGCCTCTGCCTCCCGGG 318

Qy       1317 TTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACC 1372
              ||||||  |||||||||  ||||||||||| ||||||||  ||| ||| || | |   |
Db        317 TTCAAGTGATTCTCCTGCCTCAGCTTCCCGAGTAGCTGGGGTTATAGGTGCCTCCTACCA 258

Qy       1373 ACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTG 1432
                |  | |||||||||||||||||||||||||||| ||||| ||||| || ||||||| |
Db        257 CGCCCGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGCTGGCCAGGCTGG 198

Qy       1433 CCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTAC 1492
              ||||||||||  || ||||||||||  |||||||| ||||||||||  ||||||||||
Db        197 TCTTGAACTCCCAACCTCAGGTGATCTGCCCACCTCGGCCTCCCAAAGTGCTGGGATTAC 138

Qy       1493 AGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCA 1526
              |||| |||||||||| ||||||||   ||  |||
Db        137 AGGCGTGAGCCACCGCGCCCAGCCCTCTTCTTCA 104


RESULT 41
HSM801883
LOCUS       HSM801883               3110 bp    mRNA    linear   HTC 22-SEP-2004
DEFINITION  Homo sapiens mRNA; cDNA DKFZp586G2219 (from clone DKFZp586G2219).
ACCESSION   AL136917
VERSION     AL136917.1  GI:12053328
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3110)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Agowa (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp586G2219) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp586G2219
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3110
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp586G2219"
                     /db_xref="taxon:9606"
                     /clone="DKFZp586G2219"
                     /tissue_type="uterus"
                     /clone_lib="586 (synonym: hute1). Vector pSport1; host
                     DH10B; sites NotI + SalI/MluI"
                     /dev_stage="adult"
                     /note="tuftelin 1"
     gene            1. .3110
                     /gene="DKFZp586G2219"
     CDS             51. .1223
                     /gene="DKFZp586G2219"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB66851.1"
                     /db_xref="GI:12053329"
                     /db_xref="GOA:Q9NNX1"
                     /db_xref="UniProt/Swiss-Prot:Q9NNX1"
                     /translation="MNGTRNWCTLVDVHPEDQAAGSVDILRLTLQGELTGDELEHIAQ
                     KAGRKTYAMVSSHSAGHSLASELVESHDGHEEIIKVYLKGRSGDKMIHEKNINQLKSE
                     VQYIQEARNCLQKLREDISSKLDRNLGDSLHRQEIQVVLEKPNGFSQSPTALYSSPPE
                     VDTCINEDVESLRKTVQDLLAKLQEAKRQHQSDCVAFEVTLSRYQREAEQSNVALQRE
                     EDRVEQKEAEVGELQRRLLGMETEHQALLAKVREGEVALEELRSNNADCQAEREKAAT
                     LEKEVAGLREKIHHLDDMLKSQQRKVRQMIEQLQNSKAVIQSKDATIQELKEKIAYLE
                     AENLEMHDRMEHLIEKQISHGNFSTQARAKTENPGSIRISKPPSPKPMPVIRVVET"

ORIGIN

  Query Match             9.8%;  Score 196.2;  DB 4;  Length 3110;
  Best Local Similarity   69.9%;  Pred. No. 9e-23;
  Matches  293;  Conservative    0;  Mismatches  123;  Indels    3;  Gaps    2;

Qy       1193 GTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCAC 1252
             | |   | || ||| |||       | | ||  ||||| |||||| ||| ||||||||||||
Db       2425 GCAGGTCATCATGGTTTTCTTTTTTTATTGTTTTTTTTTTTTTCTGAGACAGAGTCTCAC 2484

Qy       1253 TTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCC 1312
             | |||||||||||||||||| ||||||| | | |||| ||||||||||||||||||||||
Db       2485 TCTGTCACCCAGGCTGGAGTGCAGTGGC-GCGATCTCAGCTCACTGCAACCTCTGCCTCC 2543

Qy       1313 CAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCA 1370
             | | ||  |||||||||| |||   ||||| |||||| ||||||||| |||||| |
Db       2544 CGGATTTAAGCGATTCTTCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGTGTGCCAC 2603

Qy       1371 CCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
             |    | || || |||||||||||||||||||||||   | ||||| ||||   || | |||||
Db       2604 CACGCCTGGCTGATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATACTGGCTAGGCT 2663

Qy       1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490
              | || ||| ||||||| |||||||||||||||||||||| || ||||||||  || |||||
Db       2664 GGTCTCGAATTCCTGACCTCAGGTGATCCACCCACCTCGGCTTCCCAAAGTGCTAGGATT 2723

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCC 1550
             | |||| ||||| || |||||| |||  |  | |        |      | || ||  ||||
Db       2724 ATAGGCTTGAGCTACTGTGCCCGGCCCATGGTGTTTTTCTTTAGGGCTCTTCCTACAGCC 2783

Qy       1551 TTCATTTCTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAA 1609
              || |    ||  |    ||     | |      ||    | || ||  ||   |||||

Db       2784 TTGAGAAGTAGATAGGCATCAGAGTATGGTACTATAGGAATCAGAAAAATTCAAAACAA 2842


RESULT 42
AQ885682
LOCUS       AQ885682                 510 bp    DNA     linear   GSS 09-NOV-1999
DEFINITION  HS_5525_A2_G06_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=9293 Col=12 Row=M, genomic survey sequence.
ACCESSION   AQ885682
VERSION     AQ885682.1  GI:6317149
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 9293  row: M  column: 12
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 510.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=9293 Col=12 Row=M"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"

ORIGIN

  Query Match             9.8%;  Score 196;  DB 9;  Length 510;
  Best Local Similarity   78.0%;  Pred. No. 1.5e-22;
  Matches  273;  Conservative    0;  Mismatches   72;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             ||||||  || || |||||||||| | ||||||||  || ||||||||||||||||||  ||
Db        104 TCTTTTCTTTNTTCTTTTTTTTGAGAGAGAGTCTCGCTCTGTCACCCAGGCTGGAGTGCA 163

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
              ||||   | |||| |||||||||||||||||| |||||||| ||||||| |||||||||
Db        164 ATGGC-ACGATCTCAGCTCACTGCAACCTCTGTCTCCCAGGCTCAAGCAATTCTCCTGCC 222

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTA 1391
              ||||| |||||| |||||||||||||||| ||  |||   |   | || ||||||||||||
Db        223 TCAGCCTCCCGAGTAGCTGGGATTACAGACGTGTGCCACCATACTTGGCTAATTTTTGTA 282

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
```

```
                ||||||||||||||| | ||||| ||||| || ||||||| | ||  |||||||||| |||
Db        283 TTTTTAGTAGAGACGGNGTTTCACCATGGTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 342

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
                ||||| || ||| |||| |||||| ||||||||||||||||||||||| ||||
Db        343 TGTGATTTGCCTGCCTTGGCCTTCCAAAGTGTTGGGATTACAGGCATGAGCCACCATGCC 402

Qy       1512 CAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTAT 1561
                ||  || |||| ||   || | | ||| ||  | || |    |||| | |
Db        403 CACGCTTTTTTTTCTTTTCTTTTCTTTTTTGTTAAATCAACTATTTGTTT 452


RESULT 43
AQ379603
LOCUS       AQ379603                724 bp    DNA     linear   GSS 20-MAY-1999
DEFINITION  RPCI11-163F3.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-163F3,
            genomic survey sequence.
ACCESSION   AQ379603
VERSION     AQ379603.1  GI:4350626
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-163F3.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7562330"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-163F3"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             9.8%;  Score 196;  DB 9;  Length 724;
  Best Local Similarity   74.4%;  Pred. No. 1.4e-22;
  Matches  287;  Conservative    0;  Mismatches   95;  Indels    4;  Gaps    3;

Qy       1206 TTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGG 1265
                || ||  |  |  |||| ||| || ||||||| || |||||||| || |||| ||||||
Db        324 TTGTCATATTTGGTTTTTGTTTGTTTGTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGG 383

Qy       1266 CTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGA 1325
                ||||||| |||||||     | | ||||||||||||||||  ||||||||||||||| ||
Db        384 CTGGAGTGCAGTGGC-ACAATGTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGA 442

Qy       1326 TTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCA-CCATGCCTA 1382
                ||||||||  ||||| |||| | |||||||||||| ||||| |   |||||  |||| |||
Db        443 TTCTCCTGACTCAGCCTCCCAAGTAGCTGGGATTTCAGGCACCACCACCATGCATGGCTA 502

Qy       1383 ATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTC 1442
                |||||||||||||||||||||||||||||||| ||||||||| ||||||| | |||||||||
Db        503 ATTTTTGTATTTTTAGTAGAGACAGAGTTTTGCCATGTTGGCCAGGCTGGTCTTGAACTC 562

Qy       1443 CTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGC 1502
                ||||| ||| ||||||  |||||||| ||||||||||  ||| ||||||||||| |||||
Db        563 CTGACCTCAAGTGATCTGCCCACCTCTGCCTCCCAAAGTGCTGAGATTACAGGCTTGAGC 622

Qy       1503 CACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATC 1562
                |||| || |||||  ||||        |  ||| |   |||        | |   ||||
Db        623 CACCATGTCCAGCTGGTTTGTCATATTTGTGTTTGAATTTGGCTCTACGTGGATTCTGCG 682

Qy       1563 TTCTTATTCATTTTGGAAGTAGTACA 1588
                |     || | | |   ||| | | |
Db        683 TAAGATTTAAGTGTCTTAGTTGGAAA 708


RESULT 44
AQ075668
LOCUS       AQ075668                600 bp    DNA     linear   GSS 20-AUG-1998
DEFINITION  CIT-HSP-2367I3.TF CIT-HSP Homo sapiens genomic clone 2367I3,
            genomic survey sequence.
ACCESSION   AQ075668
VERSION     AQ075668.1  GI:3437324
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2367I3.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2367I3"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             9.8%;  Score 195.8;  DB 9;  Length 600;
  Best Local Similarity   80.0%;  Pred. No. 1.6e-22;
  Matches  268;  Conservative    0;  Mismatches   62;  Indels    5;  Gaps    3;

Qy       1194 TATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACT 1253
                | |||  || | |||  |    ||||||| ||||| |||||||||| ||||||||||||
```

```
Db        242 TCTTTTCTTTTCTTTCTTTCTTTCTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCACT 301

Qy       1254 TTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCC 1313
              |||||||| |||||||| || ||| ||  |||| ||||||||||||||||||||||
Db        302 CTGTCACCCGGGCTGGAGTGCACTGG-TGCCGTCTTGGCTCACTGCAACCTCTGCCTCCG 360

Qy       1314 AGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCAC 1371
              |||||||| |||||||||  ||||| ||| || ||||||||||||||||   | |
Db        361 GGGTTCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGTACCA 420

Qy       1372 CACCA--TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGC 1429
              |  ||  || ||||||||||||||||||||||||||| | ||||| ||||||||  |||||
Db        421 CCACACCTGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGTCAGGC 480

Qy       1430 TTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGAT 1489
              | | ||||||||||||||| ||| ||||| |||||||||  |||||||||| |||||||
Db        481 TGGTCTTGAACTCCTGACCTCAAGTGATTCACCCACCTTGGCCTCCCAAAGTGCTGGGAT 540

Qy       1490 TACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCT 1524
              | ||||| |||||||| |  ||  | |  || | |
Db        541 TGCAGGCGTGAGCCACTGCACCTGGTCCCTTCTTT 575


RESULT 45
CR859293
LOCUS       CR859293               3015 bp    mRNA    linear   HTC 12-NOV-2004
DEFINITION  Pongo pygmaeus mRNA; cDNA DKFZp468N155 (from clone DKFZp468N155).
ACCESSION   CR859293
VERSION     CR859293.1  GI:55729479
KEYWORDS    HTC.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pongo.
REFERENCE   1  (bases 1 to 3015)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp468N155) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp468N155
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3015
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp468N155"
                     /tissue_type="heart"
                     /clone_lib="468 (synonym: phrt1). Vector pSport1_Sfi; host
                     DH10B; sites SfilA + SfilB"
                     /dev_stage="adult"
                     /note="hypothetical protein (Homo sapiens)"
     gene            1. .3015
                     /gene="DKFZp468N155"
     CDS             95. .1996
                     /gene="DKFZp468N155"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH91471.1"
                     /db_xref="GI:55729480"
                     /translation="MESGPKMLAPICLVENNNEQLLVNQQAIQILEKISQPVVVVAIV
                     GLYRTGKSYLMNHLAGQNHGFPLGSTVQSETKGIWMWCVPHPSKPNHTLVLLDTEGLG
                     DVEKGDPKNDSWIFALAVLLCSTFIYNSMSTINHQALEQLHYVTELTELIKAKSSPRP
                     DGVDDSTEFVSFFPDFIWTVRDFTLELKLNGHPITEDEYLENALKLIQGNNPRVQTSN
                     LPRECIRRFFPKRKCFIFDRPTNDKDLLANIEKVSEKQLDPKFQEQTNIFSSYIFTHA
                     RTKTLREGIIVTGNRLGTLAVTYVEAVNSGAVPCLENAVITLAQRENSAAVQRAADYY
                     SQQMAQRVKFPTDTLQELLDMHAACEREAIAIFMEHSFKDENQEFQKKFMETTMNKKG
                     DFLLQNEESSVQYCQAKLNELSKGLMESISAGSFSVPGGHKLYMETKERIEQDYWQVP
                     RKGVKAKEVFQRFLESQVVIEESILQSDKALTDREKAVAVDRAKKEAAEKEQELLKQK
                     LQEQQQQMEAQDKSLKENIAQLKQKLQMEREQLLREQIMMLEHTQKVQNDWLHEGFKK
                     KYEEMNAEISQFKRMIDITKNDDTPWIARTLDKLADELTAVLSAPAKLIGHGVKGVSS
                     LFKKHKLPF"
ORIGIN

  Query Match             9.8%;  Score 195.8;  DB 4;  Length 3015;
  Best Local Similarity   70.0%;  Pred. No. 1.1e-22;
  Matches  292;  Conservative    0;  Mismatches  122;  Indels    3;  Gaps    2;

Qy       1102 CATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTTAGCCATGTGTGACAAAAGAGA 1161
              | | |||| || | ||  |||  | |||||| ||| |  |           |       ||
Db       1969 CTTTAAAAAGCATAAGCTCCCCTTTTAAGGATATTATAGATTGTACATATATGCTTTGGA 2028

Qy       1162 GGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTT 1221
                 ||| |    |    || |      | |     | | | |||| | | |  |||
Db       2029 CTATTTTTGATCTGTATGTTTTTCATTTTCATTCAGCAAGTTTTTTTTGTTTGTTTGTTT 2088

Qy       1222 TATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCT 1281
              | |||| |||||| | | |||||||||| ||| |||| || ||||||||||||||||| |
Db       2089 TGTTTTGTTTTTTTTATCAGACAGAGTCTTACTCTGTCGCCTAGGCTGGAGTACAGTGG-T 2147

Qy       1282 GTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGTCAGCTTC 1341
              |    |||| |||||||||||| |||||||||| |||||||| ||||| |||| |     |
Db       2148 GCAATCTCAGCTCACTGCAACTTCTGCCTCCCGGGTTCAAGAGATTCACCTGCCTCAGCC 2207

Qy       1342 CCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGT 1399
              || | ||||||||||||| ||| |||  |  |  ||  |  ||||||||||||||||||
Db       2208 CCTAGTAGCTGGGATTATAGGTGCACACCACCACACCCAGCTAATTTTTGTATTTTTAGT 2267

Qy       1400 AGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCC 1459
              |||||  | ||||| |||||||| ||||||| | || |||||  |||| |||  ||||||
Db       2268 AGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTTTTGACCTCAAATGATCC 2327

Qy       1460 ACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCC 1516
              |||| ||||||||||||||  |||||||||||||||||||||||||  ||||||||
Db       2328 ACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGCCCAGCC 2384


Search completed: January 22, 2006, 21:09:07
Job time : 5428.54 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 01:00:32 ; Search time 6808.48 Seconds
(without alignments)
16706.187 Million cell updates/sec

Title: US-09-728-552A-3_COPY_19000_21000
Perfect score: 2001
Sequence: 1 aatgtttggatttcaaaggt.....cagggacatgggtgcatagc 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	40917	6	AX033911	AX033911 Sequence
	2	2001	100.0	80622	8	AF222855	AF222855 Homo sapi
	3	1974.2	98.7	41008	6	AX033912	AX033912 Sequence
	4	1974.2	98.7	69058	8	AF222856	AF222856 Homo sapi
	5	1974.2	98.7	80155	8	AF042484	AF042484 Homo sapi
	6	1974.2	98.7	80202	8	AF222854	AF222854 Homo sapi
	7	1974.2	98.7	112442	8	AL355340	AL355340 Human DNA
c	8	1974.2	98.7	176432	14	AL391648	AL391648 Homo sapi
	9	670.4	33.5	71032	14	AC084084	AC084084 Homo sapi
c	10	218	10.9	27423	8	AP006290	AP006290 Homo sapi
c	11	218	10.9	133315	14	AC136997	AC136997 Homo sapi
c	12	218	10.9	161953	8	AC130304	AC130304 Homo sapi
c	13	216.6	10.8	98340	8	AL353716	AL353716 Human DNA
	14	215.8	10.8	56330	14	AL353694_3	Continuation (4 of
	15	215.2	10.8	165007	8	AC019047	AC019047 Homo sapi
	16	215	10.7	172296	14	AP002889	AP002889 Homo sapi
c	17	214	10.7	89323	8	AP000221	AP000221 Homo sapi
c	18	214	10.7	100000	8	AP000085	AP000085 Homo sapi
c	19	214	10.7	100000	8	AP000137	AP000137 Homo sapi
c	20	214	10.7	340000	8	AP001693	AP001693 Homo sapi
c	21	213.8	10.7	164857	8	AC110054	AC110054 Homo sapi
c	22	213.6	10.7	72048	6	AR659559	AR659559 Sequence
c	23	213.6	10.7	72048	6	AR659643	AR659643 Sequence
c	24	213.6	10.7	168751	14	AC068969	AC068969 Homo sapi
c	25	213.6	10.7	170883	8	AC067805	AC067805 Homo sapi
	26	213.4	10.7	175752	14	AC091721	AC091721 Pan trogl
c	27	213.4	10.7	175808	8	AC146040	AC146040 Pan trogl
c	28	213.2	10.7	42673	8	AC138517	AC138517 Homo sapi
	29	212.8	10.6	179836	8	AL590822	AL590822 Human DNA
	30	212.8	10.6	213432	14	AC068198	AC068198 Homo sapi
	31	212	10.6	130080	8	AP003160	AP003160 Homo sapi
c	32	212	10.6	163848	8	HS1056L3	AL031727 Human DNA
	33	211.8	10.6	70446	8	AC107299	AC107299 Homo sapi
	34	211.8	10.6	118135	8	AC117477	AC117477 Homo sapi
c	35	211.8	10.6	181505	14	AC025415	AC025415 Homo sapi
	36	211.6	10.6	190264	14	AC161624	AC161624 Pan trogl
	37	211.6	10.6	226170	8	AC008761	AC008761 Homo sapi
c	38	211	10.5	44553	8	D88268	D88268 Homo sapien
c	39	211	10.5	178000	8	AP002893	AP002893 Homo sapi
c	40	210.8	10.5	87071	8	CR589878	CR589878 Human DNA
c	41	210.8	10.5	107537	8	CR954268	CR954268 Human DNA
	42	210.8	10.5	138165	8	CR936851	CR936851 Human DNA
c	43	210.8	10.5	164033	8	CR938745	CR938745 Human DNA
c	44	210.8	10.5	175639	8	CR936885	CR936885 Human DNA
c	45	210.8	10.5	207664	8	AC157497	AC157497 Pan trogl
c	46	209.8	10.5	159512	14	AC144369	AC144369 Papio anu
	47	209.8	10.5	166336	8	AL445467	AL445467 Human DNA
c	48	209.8	10.5	175339	8	AC005772	AC005772 Homo sapi
	49	209.6	10.5	61505	8	AL512330	AL512330 Human DNA
c	50	209.6	10.5	162774	14	AC034138	AC034138 Homo sapi
	51	209.6	10.5	173149	14	AL391218	AL391218 Homo sapi
c	52	209.4	10.5	70248	14	AC103795	AC103795 Homo sapi
	53	209.4	10.5	149466	8	HS427A4	Z98049 Human DNA s
c	54	209.4	10.5	182329	8	AP001646	AP001646 Homo sapi
	55	209.2	10.5	145705	8	AC090255	AC090255 Homo sapi
c	56	209.2	10.5	158458	8	AC007395	AC007395 Homo sapi
	57	209.2	10.5	170394	14	AC024371	AC024371 Homo sapi
	58	209	10.4	100732	8	AL133347	AL133347 Human DNA
	59	209	10.4	107731	8	AC005476	AC005476 Homo sapi
c	60	208.8	10.4	137143	8	AL451136	AL451136 Human DNA
c	61	208.8	10.4	160342	14	AC025492	AC025492 Homo sapi
c	62	208.8	10.4	165434	14	AC087505	AC087505 Homo sapi
c	63	208.8	10.4	165546	14	AC069217	AC069217 Homo sapi
c	64	208.8	10.4	190712	14	AC165405	AC165405 Papio anu
	65	208.6	10.4	106008	8	HS5O6	AL020993 Human DNA
	66	208.6	10.4	155729	8	CNS01DWN	AL137785 Human chr
	67	208.6	10.4	178369	14	CT005241	CT005241 Pan trogl
c	68	208.6	10.4	192097	8	AC008743	AC008743 Homo sapi
c	69	208.6	10.4	196015	14	AC150024	AC150024 Papio anu
	70	208.4	10.4	6280	8	HSM805888	BX537795 Homo sapi
c	71	208.4	10.4	137289	8	AL161909	AL161909 Human DNA
c	72	208.4	10.4	148151	8	AL513533	AL513533 Human DNA
c	73	208.4	10.4	154323	14	AC026033	AC026033 Homo sapi
c	74	208.4	10.4	159367	14	AC044812	AC044812 Homo sapi
	75	208.4	10.4	164883	8	AC090282	AC090282 Homo sapi
	76	208.4	10.4	183113	8	HS243E7	AL022323 Human DNA
	77	208.4	10.4	187990	14	AL590805	AL590805 Homo sapi
	78	208.2	10.4	46919	8	AF165141	AF165141 Homo sapi
c	79	208.2	10.4	73463	8	AL161785	AL161785 Human DNA
	80	208.2	10.4	121628	8	AC117517	AC117517 Homo sapi
	81	208.2	10.4	171951	8	CNS05TDA	AL356020 Human chr
	82	208.2	10.4	192324	8	AC096920	AC096920 Homo sapi
c	83	208.2	10.4	213447	14	AC083833	AC083833 Homo sapi
	84	208	10.4	87034	8	AC004587	AC004587 Human Chr
	85	208	10.4	88955	8	AY442342	AY442342 Homo sapi
c	86	208	10.4	143796	14	AC006411	AC006411 Homo sapi
c	87	208	10.4	170245	8	HS109F14	AL022721 Human DNA
c	88	208	10.4	173701	8	AC018946	AC018946 Homo sapi
c	89	208	10.4	223972	14	AC027478	AC027478 Homo sapi
	90	207.6	10.4	71179	8	AL136110	AL136110 Human DNA
	91	207.6	10.4	108906	8	AC022088	AC022088 Homo sapi

	92	207.6	10.4	122564	8	HSJ899B16	AL050338 Human DNA
	93	207.6	10.4	145588	8	AC010000	AC010000 Homo sapi
c	94	207.6	10.4	162139	8	AC005682	AC005682 Homo sapi
c	95	207.6	10.4	164018	14	AC074242	AC074242 Homo sapi
	96	207.6	10.4	192441	14	AC073404	AC073404 Homo sapi
	97	207.6	10.4	208066	8	AC146483	AC146483 Pan trogl
	98	207.4	10.4	49616	8	AL365267	AL365267 Human DNA
	99	207.4	10.4	159266	14	AC026097	AC026097 Homo sapi
c	100	207.2	10.4	106377	8	AC004132	AC004132 Homo sapi

ALIGNMENTS

RESULT 1
AX033911
LOCUS AX033911 40917 bp DNA linear PAT 21-SEP-2000
DEFINITION Sequence 3 from Patent WO9851790.
ACCESSION AX033911
VERSION AX033911.1 GI:10280479
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified
unclassified sequences.
REFERENCE 1
AUTHORS Cancilla,M.R., Choo,K.H. and Du,S.D.
TITLE A novel nucleic acid molecule
JOURNAL Patent: WO 9851790-A 3 19-NOV-1998;
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES Location/Qualifiers
source 1. .40917
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN

Query Match 100.0%; Score 2001; DB 6; Length 40917;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19000 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 19059

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19060 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 19119

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19120 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 19179

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19180 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 19239

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19240 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 19299

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19300 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19359

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19360 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19419

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19420 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19479

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19480 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19539

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19540 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19599

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19600 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19659

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19660 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19719

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19720 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19779

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTTCAGGTGGGATGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19780 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTTCAGGTGGGATGG 19839

Qy        841 TAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGCT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19840 TAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGCT 19899

Qy        901 CTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19900 CTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCCC 19959

Qy        961 GTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19960 GTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTAG 20019

Qy       1021 AATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20020 AATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTAT 20079

Qy       1081 AGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20080 AGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTT 20139

Qy       1141 AGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20140 AGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGA 20199

Qy       1201 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20200 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCAC 20259

Qy       1261 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20260 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 20319

Qy       1321 AGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20320 AGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCC 20379

Qy       1381 TAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20380 TAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAAC 20439

Qy       1441 TCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20440 TCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGA 20499

Qy       1501 GCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20500 GCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTA 20559

Qy       1561 TCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTT 1620
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20560 TCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTT 20619

Qy       1621 TGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTT 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20620 TGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTT 20679

Qy       1681 ATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTAT 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20680 ATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTAT 20739

Qy       1741 TGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTAAG 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20740 TGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTAAG 20799

Qy       1801 TGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTGAC 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20800 TGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTGAC 20859

Qy       1861 TGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACATGC 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20860 TGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACATGC 20919

Qy       1921 TGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACGCA 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20920 TGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACGCA 20979

Qy       1981 ACAGGGACATGGGTGCATAGC 2001
                |||||||||||||||||||||
Db      20980 ACAGGGACATGGGTGCATAGC 21000
```

```
RESULT 2
AF222855
LOCUS       AF222855               80622 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222855
VERSION     AF222855.1  GI:9246845
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80622)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 80622)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Genomic sequence from human 10q25.2, clonelib=HC
FEATURES             Location/Qualifiers
     source          1. .80622
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="HC"
     repeat_region   complement(3. .176)
                     /rpt_family="Alu"
     repeat_region   complement(979. .3564)
                     /rpt_family="Tigger1"
     repeat_region   2967. .3094
                     /rpt_family="Alu"
     repeat_region   3664. .3934
                     /rpt_family="Alu"
     repeat_region   complement(4920. .5181)
                     /rpt_family="Alu"
     repeat_region   complement(6287. .6579)
                     /rpt_family="Alu"
     repeat_region   6625. .6966
                     /rpt_family="MLT2B2"
     repeat_region   complement(9215. .9490)
                     /rpt_family="Alu"
     repeat_region   complement(11280. .11565)
                     /rpt_family="Alu"
     repeat_region   complement(12428. .12675)
                     /rpt_family="Alu"
     repeat_region   complement(14304. .14583)
                     /rpt_family="Alu"
     repeat_region   17503. .17794
                     /rpt_family="Alu"
     repeat_region   complement(20250. .20533)
                     /rpt_family="Alu"
     repeat_region   complement(23031. .23323)
                     /rpt_family="Alu"
     repeat_region   complement(24301. .24479)
                     /rpt_family="MIR"
     repeat_region   24951. .25250
                     /rpt_family="Alu"
     repeat_region   complement(25987. .26342)
                     /rpt_family="THE1"
     repeat_region   27097. .27335
                     /rpt_family="MER2"
     repeat_region   27538. .27814
                     /rpt_family="Alu"
     repeat_region   27891. .28161
                     /rpt_family="Alu"
     repeat_region   29063. .29334
                     /rpt_family="Alu"
     repeat_region   complement(31025. .31093)
                     /rpt_family="MLT1"
     repeat_region   complement(31817. .32352)
                     /rpt_family="MER1"
     repeat_region   complement(32489. .32607)
                     /rpt_family="MLT1"
     repeat_region   33039. .33316
                     /rpt_family="Alu"
     repeat_region   33469. .33760
                     /rpt_family="Alu"
     repeat_region   complement(34527. .34590)
                     /rpt_family="MIR"
     repeat_region   34689. .34995
                     /rpt_family="Alu"
     repeat_region   complement(36349. .36437)
                     /rpt_family="MIR"
     repeat_region   complement(38796. .39079)
                     /rpt_family="Alu"
     repeat_region   39441. .39693
                     /rpt_family="Alu"
     repeat_region   complement(43176. .43237)
                     /rpt_family="MIR"
     repeat_region   complement(45203. .45478)
                     /rpt_family="Alu"
     repeat_region   45892. .46025
                     /rpt_family="MIR"
     repeat_region   complement(46618. .46851)
                     /rpt_family="Alu"
     repeat_region   48623. .48939
                     /rpt_family="Alu"
     repeat_region   complement(49048. .50698)
                     /rpt_family="L1"
     repeat_region   complement(50710. .50972)
                     /rpt_family="Alu"
     repeat_region   complement(51031. .51192)
                     /rpt_family="L1"
     repeat_region   complement(51450. .51739)
```

```
                     /rpt_family="Alu"
     repeat_region   complement(53270. .53562)
                     /rpt_family="Alu"
     repeat_region   54261. .54808
                     /rpt_family="LTR12"
     repeat_region   59460. .60054
                     /rpt_family="LTR12"
     repeat_region   60349. .60633
                     /rpt_family="Alu"
     repeat_region   complement(60669. .60801)
                     /rpt_family="MIR"
     repeat_region   complement(61684. .61852)
                     /rpt_family="MER5"
     repeat_region   63466. .63753
                     /rpt_family="Alu"
     repeat_region   64063. .64228
                     /rpt_family="MER5"
     repeat_region   64446. .64521
                     /rpt_family="MIR"
     repeat_region   65794. .65883
                     /rpt_family="MER5"
     repeat_region   66772. .66951
                     /rpt_family="MER20"
     repeat_region   complement(67895. .68416)
                     /rpt_family="MER41"
     repeat_region   complement(73192. .73470)
                     /rpt_family="Alu"
     repeat_region   74086. .74367
                     /rpt_family="Alu"
     repeat_region   75916. .76183
                     /rpt_family="Alu"
     repeat_region   complement(77771. .78059)
                     /rpt_family="Alu"
ORIGIN

  Query Match             100.0%;  Score 2001;  DB 8;  Length 80622;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19027 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 19086

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19087 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 19146

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19147 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 19206

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19207 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 19266

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19267 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 19326

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19327 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19386

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19387 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19446

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19447 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19506

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      19507 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19566

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19567 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19626

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19627 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19686

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19687 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19746

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19747 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19806

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTTCAGGTGGGATGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19807 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTTCAGGTGGGATGG 19866

Qy        841 TAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGCT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19867 TAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGCT 19926

Qy        901 CTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19927 CTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCCC 19986

Qy        961 GTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19987 GTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTAG 20046

Qy       1021 AATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20047 AATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTAT 20106

Qy       1081 AGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20107 AGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTT 20166

Qy       1141 AGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20167 AGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGA 20226

Qy       1201 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20227 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCAC 20286

Qy       1261 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20287 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 20346

Qy       1321 AGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20347 AGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCC 20406

Qy       1381 TAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20407 TAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAAC 20466

Qy       1441 TCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20467 TCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGA 20526

Qy       1501 GCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20527 GCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTA 20586

Qy       1561 TCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20587 TCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTT 20646
```

```
Qy       1621 TGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTT 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20647 TGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTT 20706

Qy       1681 ATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTAT 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20707 ATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTAT 20766

Qy       1741 TGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTAAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20767 TGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTAAG 20826

Qy       1801 TGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTGAC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20827 TGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTGAC 20886

Qy       1861 TGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACATGC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20887 TGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACATGC 20946

Qy       1921 TGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACGCA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20947 TGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACGCA 21006

Qy       1981 ACAGGGACATGGGTGCATAGC 2001
             |||||||||||||||||||||
Db      21007 ACAGGGACATGGGTGCATAGC 21027
```

RESULT 3
AX033912
LOCUS AX033912 41008 bp DNA linear PAT 21-SEP-2000
DEFINITION Sequence 4 from Patent WO9851790.
ACCESSION AX033912
VERSION AX033912.1 GI:10280480
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified
unclassified sequences.
REFERENCE 1
AUTHORS Cancilla,M.R., Choo,K.H. and Du,S.D.
TITLE A novel nucleic acid molecule
JOURNAL Patent: WO 9851790-A 4 19-NOV-1998;
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES Location/Qualifiers
source 1. .41008
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN

Query Match 98.7%; Score 1974.2; DB 6; Length 41008;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1998; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```
Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18736 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 18795

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18796 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 18855

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18856 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 18915

Qy        181 AAAGCCAAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18916 AAAGCCAAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 18975

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18976 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 19035

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      19036 CATATGCAGGAGTGTCACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19095

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19096 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19155

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19156 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19215

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19216 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19275

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19276 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19335

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19336 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19395

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19396 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19455

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19456 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19515

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTT-TCAGGTGGGATG 839
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      19516 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTCTCAGGTGGGATG 19575

Qy        840 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19576 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 19635

Qy        900 TCTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCC 959
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      19636 TCTGACTGGGGACCCTAAATATGCAAAAGTTAATAGGCTCTTCATGCAGAATATGAACCC 19695

Qy        960 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19696 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 19755

Qy       1020 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19756 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 19815

Qy       1080 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19816 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 19875

Qy       1140 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19876 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 19935

Qy       1200 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19936 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 19995

Qy       1260 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19996 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 20055

Qy       1320 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 1379
```

```
Db      20056 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 20115

Qy       1380 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 1439

Db      20116 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 20175

Qy       1440 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 1499

Db      20176 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 20235

Qy       1500 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA-TTTGTTTCTGAAGCCTTCATTTC 1558

Db      20236 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTTGTTTCTGAAGCCTTCATTTC 20295

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618

Db      20296 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 20355

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678

Db      20356 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 20415

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738

Db      20416 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 20475

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798

Db      20476 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 20535

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858

Db      20536 ATTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 20595

Qy       1859 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 1918

Db      20596 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 20655

Qy       1919 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 1978

Db      20656 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 20715

Qy       1979 CAACAGGGACATGGGTGCATAGC 2001

Db      20716 CAACAGGGACATGGGTGCATAGC 20738
```

```
RESULT 4
AF222856
LOCUS       AF222856               69058 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone PnC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222856
VERSION     AF222856.1  GI:9246846
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 69058)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 69058)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Human genomic sequence from 10q25.2, clonelib=PnC
FEATURES             Location/Qualifiers
     source          1. .69058
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="PnC"
     repeat_region   complement(598. .3183)
                     /rpt_family="Tigger1"
     repeat_region   2587. .2715
                     /rpt_family="Alu"
     repeat_region   3283. .3553
                     /rpt_family="Alu"
     repeat_region   complement(4548. .4809)
                     /rpt_family="Alu"
     repeat_region   complement(5919. .6211)
                     /rpt_family="Alu"
     repeat_region   6257. .6599
                     /rpt_family="MLT2B2"
     repeat_region   complement(8848. .9123)
                     /rpt_family="Alu"
     repeat_region   complement(10913. .11198)
                     /rpt_family="Alu"
     repeat_region   complement(12062. .12313)
                     /rpt_family="Alu"
     repeat_region   complement(13957. .14236)
                     /rpt_family="Alu"
     repeat_region   17133. .17424
                     /rpt_family="Alu"
     repeat_region   complement(19882. .20197)
                     /rpt_family="Alu"
     repeat_region   complement(22664. .22956)
                     /rpt_family="Alu"
     repeat_region   complement(23935. .24113)
                     /rpt_family="MIR"
     repeat_region   24585. .24884
                     /rpt_family="Alu"
     repeat_region   complement(25618. .25973)
                     /rpt_family="THE1"
     repeat_region   26732. .26970
                     /rpt_family="MER2"
     repeat_region   27173. .27449
                     /rpt_family="Alu"
     repeat_region   27526. .27796
                     /rpt_family="Alu"
     repeat_region   28702. .28978
                     /rpt_family="Alu"
     repeat_region   complement(30668. .30736)
                     /rpt_family="MLT1"
     repeat_region   complement(31460. .31993)
                     /rpt_family="MER1"
     repeat_region   complement(32130. .32248)
                     /rpt_family="MLT1"
     repeat_region   32685. .32961
                     /rpt_family="Alu"
     repeat_region   33113. .33403
                     /rpt_family="Alu"
     repeat_region   complement(34169. .34233)
                     /rpt_family="MIR"
     repeat_region   34332. .34639
                     /rpt_family="Alu"
     repeat_region   complement(35993. .36081)
                     /rpt_family="MIR"
     repeat_region   complement(38438. .38721)
                     /rpt_family="Alu"
     repeat_region   39081. .39333
                     /rpt_family="Alu"
     repeat_region   complement(42817. .42877)
                     /rpt_family="MIR"
     repeat_region   complement(44845. .45119)
```

```
                     /rpt_family="Alu"
     repeat_region   45532. .45665
                     /rpt_family="MIR"
     repeat_region   complement(46257. .46490)
                     /rpt_family="Alu"
     repeat_region   48262. .48578
                     /rpt_family="Alu"
     repeat_region   complement(48687. .50337)
                     /rpt_family="L1"
     repeat_region   complement(50349. .50611)
                     /rpt_family="Alu"
     repeat_region   complement(50670. .50831)
                     /rpt_family="L1"
     repeat_region   complement(51085. .51374)
                     /rpt_family="Alu"
     repeat_region   complement(52905. .53197)
                     /rpt_family="Alu"
     repeat_region   53896. .54467
                     /rpt_family="LTR12"
     repeat_region   59129. .59723
                     /rpt_family="LTR12"
     repeat_region   60018. .60302
                     /rpt_family="Alu"
     repeat_region   complement(60338. .60470)
                     /rpt_family="MIR"
     repeat_region   complement(61353. .61521)
                     /rpt_family="MER5"
     repeat_region   63135. .63416
                     /rpt_family="Alu"
     repeat_region   63724. .63889
                     /rpt_family="MER5"
     repeat_region   64107. .64182
                     /rpt_family="MIR"
     repeat_region   65459. .65548
                     /rpt_family="MER5"
     repeat_region   66439. .66618
                     /rpt_family="MER20"
     repeat_region   complement(67562. .68084)
                     /rpt_family="MER41"
ORIGIN

  Query Match             98.7%;  Score 1974.2;  DB 8;  Length 69058;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1998;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18658 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 18717

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18718 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 18777

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18778 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 18837

Qy        181 AAAGCCAAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18838 AAAGCCAAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 18897

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18898 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 18957

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18958 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19017

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19018 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19077

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19078 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19137

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19138 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19197

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19198 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19257

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19258 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19317

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19318 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19377

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19378 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19437

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTT-TCAGGTGGGATG 839
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      19438 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTCTCAGGTGGGATG 19497

Qy        840 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19498 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 19557

Qy        900 TCTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCC 959
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      19558 TCTGACTGGGGACCCTAAATATGCAAAAGTTAATAGGCTCTTCATGCAGAATATGAACCC 19617

Qy        960 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19618 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 19677

Qy       1020 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19678 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 19737

Qy       1080 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19738 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 19797

Qy       1140 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 1199
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      19798 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTCTTCAGAGAAGTATTTG 19857

Qy       1200 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19858 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 19917

Qy       1260 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19918 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 19977

Qy       1320 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19978 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 20037

Qy       1380 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20038 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 20097

Qy       1440 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20098 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 20157

Qy       1500 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA-TTTGTTTCTGAAGCCTTCATTTC 1558
```

```
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      20158 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTTGTTTCTGAAGCCTTCATTTC 20217

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20218 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 20277

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20278 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 20337

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20338 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 20397

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20398 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 20457

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858
                | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20458 ATTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 20517

Qy       1859 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 1918
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20518 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 20577

Qy       1919 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 1978
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20578 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 20637

Qy       1979 CAACAGGGACATGGGTGCATAGC 2001
                |||||||||||||||||||||||
Db      20638 CAACAGGGACATGGGTGCATAGC 20660
```

```
RESULT 5
AF042484
LOCUS       AF042484               80155 bp    DNA     linear   PRI 22-JUN-2000
DEFINITION  Homo sapiens sequence of the 80 kb core centromere protein binding
            region of the mar del(10) neocentromere corresponding to human
            chromosome 10q25.2, complete sequence.
ACCESSION   AF042484
VERSION     AF042484.1  GI:4205782
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,K.H.
  TITLE     Sequence analysis of an 80 kb human neocentromere
  JOURNAL   Hum. Mol. Genet. 8 (2), 217-227 (1999)
   PUBMED   9931329
REFERENCE   2  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville, Melbourne, Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .80155
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /note="Sequence information obtained from BAC clones
                     prepared from cell line BE2C1-18-5f"
     repeat_region   complement(join(49. .525,3628. .4545,4849. .5053))
                     /rpt_family="L1MEc"
                     /rpt_type=dispersed
     repeat_region   join(560. .597,3184. .3281)
                     /rpt_family="MER58"
                     /rpt_type=dispersed
     repeat_region   complement(join(598. .2553,2716. .3183))
                     /rpt_family="TIGGER1"
                     /rpt_type=dispersed
     repeat_region   2558. .2581
                     /standard_name="T1"
                     /note="100% homology"
                     /rpt_type=direct
                     /rpt_unit="(caggcacagtgg)2"
     repeat_region   2570. .2715
                     /rpt_family="AluSg/x"
                     /rpt_type=dispersed
     repeat_region   3021. .3050
                     /standard_name="T2"
                     /note="83.3% homology"
                     /rpt_type=direct
                     /rpt_unit="(taacaaagtg)3"
     repeat_region   3283. .3558
                     /rpt_family="AluSx"
                     /rpt_type=dispersed
     repeat_region   4173. .4211
                     /standard_name="T3"
                     /note="82.1% homology"
                     /rpt_type=direct
                     /rpt_unit="(aaaaaaataattt)3"
     repeat_region   complement(4546. .4848)
                     /rpt_family="AluSx"
                     /rpt_type=dispersed
     repeat_region   complement(5909. .6208)
                     /rpt_family="AluSg"
                     /rpt_type=dispersed
     repeat_region   join(6232. .6573,6638. .6688)
                     /rpt_family="MLT2CB"
                     /rpt_type=dispersed
     repeat_region   6598. .6645
                     /standard_name="AC24"
                     /note="100% homology"
                     /rpt_type=tandem
                     /rpt_unit="ac"
     repeat_region   6598. .6637
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ca"
     repeat_region   7091. .7134
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   7808. .7911
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   8349. .8524
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   complement(8839. .9124)
                     /rpt_family="AluJb"
                     /rpt_type=dispersed
     repeat_region   9180. .9280
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   10070. .10170
                     /note="low complexity"
                     /rpt_family="purine-rich"
     repeat_region   10097. .10136
                     /standard_name="GA1"
                     /note="82.5% homology"
                     /rpt_type=direct
                     /rpt_unit="(gaaagaaagg)4"
     repeat_region   10175. .10339
                     /rpt_family="FRAM"
                     /rpt_type=dispersed
     repeat_region   complement(10899. .11197)
                     /rpt_family="AluSx"
```

```
                /rpt_type=dispersed
misc_feature    join(11627. .11757,12434. .12512,13443. .13562)
                /note=">98% identity"
                /db_xref="dbEST:AA774571"
repeat_region   11996. .12028
                /standard_name="AT4"
                /note="100% homology"
                /rpt_type=tandem
                /rpt_unit="(attt)8"
repeat_region   complement(12010. .12312)
                /rpt_family="AluJb"
                /rpt_type=dispersed
repeat_region   12327. .12433
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(13946. .14235)
                /rpt_family="AluSq"
                /rpt_type=dispersed
repeat_region   15130. .15159
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   15164. .15753
                /standard_name="AT28"
                /note="82.1% homology"
                /rpt_type=direct
                /rpt_unit="(atgtatatatgtgtatatagacataaat)21"
repeat_region   join(15170. .15289,15410. .15612)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="ta"
repeat_region   15617. .15731
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="cata"
repeat_region   16004. .16100
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16128. .16364)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16407. .16592)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16657. .16889)
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   16914. .16936
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   17102. .17386
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   17406. .17564
                /standard_name="GA2"
                /note="81.8% homology"
                /rpt_type=direct
repeat_region   17416. .17543
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   17647. .17702
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   complement(19863. .20161)
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   complement(20842. .20913)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(21366. .21414)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   21375. .21639
                /standard_name="CT1"
                /note="75% homology"
                /rpt_type=dispersed
repeat_region   complement(21415. .21536)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="tggaa"
repeat_region   complement(21524. .21646)
                /note="low complexity"
                /rpt_family="purine-rich"
repeat_region   22409. .22646
                /rpt_family="L1M4"
                /rpt_type=dispersed
repeat_region   22560. .22604
                /standard_name="T4"
                /note="80% homology"
                /rpt_type=direct
                /rpt_unit="(aatattacaataatt)3"
```

Query Match 98.7%; Score 1974.2; DB 8; Length 80155;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1998; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```
Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18645 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 18704

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18705 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 18764

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18765 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 18824

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18825 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 18884

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18885 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 18944

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      18945 CATATGCAGGAGTGTCACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19004

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19005 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19064

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19065 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19124

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19125 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19184

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19185 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19244

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19245 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19304

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19305 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19364

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19365 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19424

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTT-TCAGGTGGGATG 839
                |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      19425 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTCTCAGGTGGGATG 19484

Qy        840 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 899
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19485 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 19544

Qy        900 TCTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCC 959
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      19545 TCTGACTGGGGACCCTAAATATGCAAAAGTTAATAGGCTCTTCATGCAGAATATGAACCC 19604

Qy        960 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 1019
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19605 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 19664

Qy       1020 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 1079
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19665 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 19724

Qy       1080 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 1139
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19725 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 19784

Qy       1140 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 1199
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19785 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 19844

Qy       1200 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 1259
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19845 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 19904

Qy       1260 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 1319
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19905 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 19964

Qy       1320 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 1379
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19965 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 20024

Qy       1380 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 1439
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20025 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 20084

Qy       1440 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 1499
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20085 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 20144

Qy       1500 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA-TTTGTTTCTGAAGCCTTCATTTC 1558
                |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db      20145 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTTGTTTCTGAAGCCTTCATTTC 20204

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20205 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 20264

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20265 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 20324

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20325 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 20384

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20385 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 20444

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858
                | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      20445 ATTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 20504

Qy       1859 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 1918
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20505 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 20564

Qy       1919 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 1978
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20565 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 20624

Qy       1979 CAACAGGGACATGGGTGCATAGC 2001
                |||||||||||||||||||||||
Db      20625 CAACAGGGACATGGGTGCATAGC 20647
```

```
RESULT 6
AF222854
LOCUS       AF222854               80202 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222854
VERSION     AF222854.1  GI:9246844
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80202)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 80202)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Human genomic sequence from 10q25.2, clonelib=NC, second release
FEATURES             Location/Qualifiers
     source          1. .80202
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="NC"
     repeat_region   complement(598. .3183)
                     /rpt_family="Tigger1"
     repeat_region   2587. .2715
                     /rpt_family="Alu"
     repeat_region   3283. .3553
                     /rpt_family="Alu"
     repeat_region   complement(4548. .4809)
                     /rpt_family="Alu"
     repeat_region   complement(5919. .6211)
                     /rpt_family="Alu"
     repeat_region   6257. .6599
                     /rpt_family="MLT2B2"
     repeat_region   complement(8848. .9123)
                     /rpt_family="Alu"
     repeat_region   complement(10913. .11198)
                     /rpt_family="Alu"
     repeat_region   complement(12062. .12313)
                     /rpt_family="Alu"
     repeat_region   complement(13957. .14236)
                     /rpt_family="Alu"
     repeat_region   17133. .17424
                     /rpt_family="Alu"
     repeat_region   complement(19882. .20197)
```

```
                     /rpt_family="Alu"
     repeat_region   complement(22664. .22956)
                     /rpt_family="Alu"
     repeat_region   complement(23935. .24113)
                     /rpt_family="MIR"
     repeat_region   24585. .24884
                     /rpt_family="Alu"
     repeat_region   complement(25618. .25973)
                     /rpt_family="THE1"
     repeat_region   26732. .26970
                     /rpt_family="MER2"
     repeat_region   27173. .27449
                     /rpt_family="Alu"
     repeat_region   27526. .27796
                     /rpt_family="Alu"
     repeat_region   28702. .28978
                     /rpt_family="Alu"
     repeat_region   complement(30668. .30736)
                     /rpt_family="MLT1"
     repeat_region   complement(31460. .31993)
                     /rpt_family="MER1"
     repeat_region   complement(32130. .32248)
                     /rpt_family="MLT1"
     repeat_region   32685. .32961
                     /rpt_family="Alu"
     repeat_region   33113. .33403
                     /rpt_family="Alu"
     repeat_region   complement(34169. .34233)
                     /rpt_family="MIR"
     repeat_region   34332. .34639
                     /rpt_family="Alu"
     repeat_region   complement(35993. .36081)
                     /rpt_family="MIR"
     repeat_region   complement(38438. .38721)
                     /rpt_family="Alu"
     repeat_region   39081. .39333
                     /rpt_family="Alu"
     repeat_region   complement(42817. .42877)
                     /rpt_family="MIR"
     repeat_region   complement(44845. .45119)
                     /rpt_family="Alu"
     repeat_region   45532. .45665
                     /rpt_family="MIR"
     repeat_region   complement(46257. .46490)
                     /rpt_family="Alu"
     repeat_region   48262. .48578
                     /rpt_family="Alu"
     repeat_region   complement(48687. .50337)
                     /rpt_family="L1"
     repeat_region   complement(50349. .50611)
                     /rpt_family="Alu"
     repeat_region   complement(50670. .50831)
                     /rpt_family="L1"
     repeat_region   complement(51085. .51374)
                     /rpt_family="Alu"
     repeat_region   complement(52905. .53197)
                     /rpt_family="Alu"
     repeat_region   53896. .54467
                     /rpt_family="LTR12"
     repeat_region   59129. .59719
                     /rpt_family="LTR12"
     repeat_region   60018. .60302
                     /rpt_family="Alu"
     repeat_region   complement(60338. .60470)
                     /rpt_family="MIR"
     repeat_region   complement(61353. .61521)
                     /rpt_family="MER5"
     repeat_region   63135. .63416
                     /rpt_family="Alu"
     repeat_region   63724. .63889
                     /rpt_family="MER5"
     repeat_region   64107. .64182
                     /rpt_family="MIR"
     repeat_region   65459. .65548
                     /rpt_family="MER5"
     repeat_region   66439. .66618
                     /rpt_family="MER20"
     repeat_region   complement(67562. .68084)
                     /rpt_family="MER41"
     repeat_region   complement(72861. .73139)
                     /rpt_family="Alu"
     repeat_region   73756. .74037
                     /rpt_family="Alu"
     repeat_region   75588. .75853
                     /rpt_family="Alu"
     repeat_region   complement(77441. .77729)
                     /rpt_family="Alu"
ORIGIN

  Query Match             98.7%;  Score 1974.2;  DB 8;  Length 80202;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1998;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18658 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 18717

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18718 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 18777

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18778 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 18837

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18838 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 18897

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18898 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 18957

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18958 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19017

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19018 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19077

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19078 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19137

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19138 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19197

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19198 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19257

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19258 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19317

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19318 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19377

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19378 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19437

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTT-TCAGGTGGGATG 839
```

```
Db      19438 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTCTCAGGTGGGATG 19497

Qy        840 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 899

Db      19498 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 19557

Qy        900 TCTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCC 959

Db      19558 TCTGACTGGGGACCCTAAATATGCAAAAGTTAATAGGCTCTTCATGCAGAATATGAACCC 19617

Qy        960 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 1019

Db      19618 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 19677

Qy       1020 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 1079

Db      19678 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 19737

Qy       1080 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 1139

Db      19738 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 19797

Qy       1140 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 1199

Db      19798 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTCTTCAGAGAAGTATTTG 19857

Qy       1200 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 1259

Db      19858 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 19917

Qy       1260 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 1319

Db      19918 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 19977

Qy       1320 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 1379

Db      19978 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 20037

Qy       1380 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 1439

Db      20038 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 20097

Qy       1440 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 1499

Db      20098 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 20157

Qy       1500 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA-TTTGTTTCTGAAGCCTTCATTTC 1558

Db      20158 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTTGTTTCTGAAGCCTTCATTTC 20217

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618

Db      20218 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 20277

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678

Db      20278 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 20337

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738

Db      20338 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 20397

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798

Db      20398 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 20457

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858

Db      20458 ATTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 20517

Qy       1859 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 1918

Db      20518 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 20577

Qy       1919 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 1978

Db      20578 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 20637

Qy       1979 CAACAGGGACATGGGTGCATAGC 2001

Db      20638 CAACAGGGACATGGGTGCATAGC 20660
```

```
RESULT 7
AL355340
LOCUS       AL355340              112442 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-383C6 on chromosome 10 Contains
            the TRUB1 gene for TruB pseudouridine (psi) synthase homolog 1 (E.
            coli), a novel gene and a CpG island, complete sequence.
ACCESSION   AL355340
VERSION     AL355340.17  GI:21436504
KEYWORDS    HTG; TRUB1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 112442)
  AUTHORS   Leongamornlert,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 17, 2002 this sequence version replaced gi:16416169.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-383C6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .112442
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIB753C06383"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-383C6"
                     /clone_lib="RPCI-11.2"
     misc_feature    2000
                     /note="Clone_right_end: RP11-106M7"
     gene            join(24099. .24445,28551. .28649,37000. .37055,45632. .45713,
                     56274. .56346,58041. .58180,60235. .60291,61029. .63577)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
```

```
     mRNA            join(24099. .24445,28551. .28649,37000. .37055,45632. .45713,
                     56274. .56346,58041. .58180,60235. .60291,61029. .63577)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
                     /product="TruB pseudouridine (psi) synthase homolog 1 (E.
                     coli)"
                     /note="match: ESTs: AI208778.1 BQ775335.1
                     match: cDNAs: AF448144.1 BC030601.1"
     CDS             join(24160. .24445,28551. .28649,37000. .37055,45632. .45713,
                     56274. .56346,58041. .58180,60235. .60291,61029. .61285)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
                     /standard_name="OTTHUMP00000020554"
                     /note="match: proteins: Q8WWH5 Q9D331"
                     /codon_start=1
                     /product="TruB pseudouridine (psi) synthase homolog 1 (E.
                     coli)"
                     /protein_id="CAH73219.1"
                     /db_xref="GI:55663184"
                     /db_xref="Genew:16060"
                     /db_xref="GOA:Q8WWH5"
                     /db_xref="InterPro:IPR002501"
                     /db_xref="InterPro:IPR004510"
                     /db_xref="UniProt/TrEMBL:Q8WWH5"
                     /translation="MAASEAAVVSSPSLKTDTSPVLETAGTVAAMAATPSARAAAAVV
                     AAAARTGSEARVSKAALATKLLSLSGVFAVHKPKGPTSAELLNRLKEKLLAEAGMPSP
                     EWTKRKKQTLKIGHGGTLDSAARGVLVVGIGSGTKMLTSMLSGSKRYTAIGELGKATD
                     TLDSTGRVTEEKPYDKITQEDIEGILQKFTGNIMQVPPLYSALKKDGQRLSTLMKRGE
                     VVEAKPARPVTVYSISLQKFQPPFFTLDVECGGGFYIRSLVSDIGKELSSCANVLELT
                     RTKQGPFTLEEHALPEDKWTIDDIAQSLEHCSSLFPAELALKKSKPESNEQVLSCEYI
                     TLNEPKREDDVIKTC"
     gene            join(24292. .24445,28551. .28649,37000. .37055,45632. .45906)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-002"
     mRNA            join(24292. .24445,28551. .28649,37000. .37055,45632. .45906)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-002"
                     /product="TruB pseudouridine (psi) synthase homolog 1 (E.
                     coli)"
                     /note="match: ESTs: AI418848.1 AI422325.1 AI640751.1
                     BE858241.1"
     polyA_signal    61455. .61460
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
     polyA_site      61479
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
     polyA_signal    63558. .63563
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
     polyA_site      63582
     gene            join(80563. .80691,81349. .81444,82381. .82499)
                     /locus_tag="RP11-383C6.2-001"
     mRNA            join(80563. .80691,81349. .81444,82381. .82499)
                     /locus_tag="RP11-383C6.2-001"
                     /product="putative novel transcript"
                     /note="match: ESTs: AA774571.1"
     misc_feature    110443
                     /note="Clone_left_end: RP11-338L11"
ORIGIN
```

Query Match 98.7%; Score 1974.2; DB 8; Length 112442;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1998; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```
Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87610 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 87669

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87670 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 87729

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87730 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 87789

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87790 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 87849

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87850 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 87909

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87910 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 87969

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87970 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 88029

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88030 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 88089

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88090 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 88149

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88150 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 88209

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88210 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 88269

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88270 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 88329

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88330 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 88389

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTT-TCAGGTGGGATG 839
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      88390 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTCTCAGGTGGGATG 88449

Qy        840 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88450 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 88509

Qy        900 TCTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCC 959
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      88510 TCTGACTGGGGACCCTAAATATGCAAAAGTTAATAGGCTCTTCATGCAGAATATGAACCC 88569

Qy        960 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88570 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 88629

Qy       1020 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88630 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 88689

Qy       1080 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88690 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 88749

Qy       1140 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 1199
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      88750 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTCTTCAGAGAAGTATTTG 88809

Qy       1200 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 1259
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88810 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 88869

Qy       1260 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 1319
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88870 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 88929

Qy       1320 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 1379
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88930 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 88989

Qy       1380 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 1439
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88990 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 89049

Qy       1440 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 1499
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89050 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 89109

Qy       1500 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA-TTTGTTTCTGAAGCCTTCATTTC 1558
                |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      89110 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTTGTTTCTGAAGCCTTCATTTC 89169

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89170 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 89229

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89230 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 89289

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89290 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 89349

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89350 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 89409

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858
                | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89410 ATTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 89469

Qy       1859 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 1918
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89470 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 89529

Qy       1919 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 1978
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89530 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 89589

Qy       1979 CAACAGGGACATGGGTGCATAGC 2001
                |||||||||||||||||||||||
Db      89590 CAACAGGGACATGGGTGCATAGC 89612
```

```
RESULT 8
AL391648/c
LOCUS       AL391648              176432 bp    DNA     linear   HTG 18-JUL-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-203A7, 9 unordered pieces.
ACCESSION   AL391648
VERSION     AL391648.12  GI:14669309
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 10, 2001 this sequence version replaced gi:14596357.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA203A7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 174005 bases at least Q40
            Consensus quality: 174806 bases at least Q30
            Consensus quality: 175273 bases at least Q20
            Insert size: 175632; sum-of-contigs
            Insert size: 155595; 14.1% error; agarose-fp
            Quality coverage: 6.85x in Q20 bases; sum-of-contigs Quality
            coverage: 8.13x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    23099: contig of 23099 bp in length
            *    23100    23199: gap of 100 bp
            *    23200    38049: contig of 14850 bp in length
            *    38050    38149: gap of 100 bp
            *    38150    65430: contig of 27281 bp in length
            *    65431    65530: gap of 100 bp
            *    65531   100606: contig of 35076 bp in length
            *   100607   100706: gap of 100 bp
            *   100707   156710: contig of 56004 bp in length
            *   156711   156810: gap of 100 bp
            *   156811   159233: contig of 2423 bp in length
            *   159234   159333: gap of 100 bp
            *   159334   171732: contig of 12399 bp in length
            *   171733   171832: gap of 100 bp
            *   171833   174095: contig of 2263 bp in length
            *   174096   174195: gap of 100 bp
            *   174196   176432: contig of 2237 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176432
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-203A7"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .23099
                     /note="assembly_fragment:01553
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    23200. .38049
                     /note="assembly_fragment:00834
                     fragment_chain:1"
     misc_feature    38150. .65430
                     /note="assembly_fragment:02112
                     fragment_chain:1"
     misc_feature    65531. .100606
                     /note="assembly_fragment:01440
                     fragment_chain:1"
     misc_feature    100707. .156710
                     /note="assembly_fragment:01209
                     fragment_chain:1"
     misc_feature    156811. .159233
                     /note="assembly_fragment:00074"
```

```
     misc_feature    159334. .171732
                     /note="assembly_fragment:01455"
     misc_feature    171833. .174095
                     /note="assembly_fragment:01550"
     misc_feature    174196. .176432
                     /note="assembly_fragment:03092"
ORIGIN

  Query Match             98.7%;  Score 1974.2;  DB 14;  Length 176432;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1998;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26028 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 25969

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25968 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 25909

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25908 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 25849

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25848 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 25789

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25788 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 25729

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25728 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 25669

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25668 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 25609

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25608 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 25549

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25548 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 25489

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25488 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 25429

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25428 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 25369

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25368 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 25309

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25308 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 25249

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTT-TCAGGTGGGATG 839
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      25248 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTCTCAGGTGGGATG 25189

Qy        840 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25188 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 25129

Qy        900 TCTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCC 959
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      25128 TCTGACTGGGGACCCTAAATATGCAAAAGTTAATAGGCTCTTCATGCAGAATATGAACCC 25069

Qy        960 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25068 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 25009

Qy       1020 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25008 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 24949

Qy       1080 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24948 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 24889

Qy       1140 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 1199
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      24888 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTCTTCAGAGAAGTATTTG 24829

Qy       1200 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCACTTTGTCA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24828 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCACTTTGTCA 24769

Qy       1260 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24768 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 24709

Qy       1320 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24708 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 24649

Qy       1380 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24648 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 24589

Qy       1440 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24588 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 24529

Qy       1500 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA-TTTGTTTCTGAAGCCTTCATTTC 1558
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      24528 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTTGTTTCTGAAGCCTTCATTTC 24469

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24468 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 24409

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24408 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 24349

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24348 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 24289

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24288 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 24229

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24228 ATTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 24169

Qy       1859 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 1918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24168 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 24109

Qy       1919 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 1978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24108 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 24049

Qy       1979 CAACAGGGACATGGGTGCATAGC 2001
              |||||||||||||||||||||||
```

```
Db      24048 CAACAGGGACATGGGTGCATAGC 24026


RESULT 9
AC084084
LOCUS       AC084084               71032 bp    DNA     linear   HTG 25-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-314C19 map 8, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC084084
VERSION     AC084084.2  GI:13446278
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 71032)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-314C19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71032)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 25, 2001 this sequence version replaced gi:10799449.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11327
                Center clone name: 314_C_19
            --------------.
            * NOTE: This record contains 85 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      720: contig of 720 bp in length
            *      721      820: gap of 100 bp
            *      821     1557: contig of 737 bp in length
            *     1558     1657: gap of 100 bp
            *     1658     2404: contig of 747 bp in length
            *     2405     2504: gap of 100 bp
            *     2505     3299: contig of 795 bp in length
            *     3300     3399: gap of 100 bp
            *     3400     4153: contig of 754 bp in length
            *     4154     4253: gap of 100 bp
            *     4254     4978: contig of 725 bp in length
            *     4979     5078: gap of 100 bp
            *     5079     5815: contig of 737 bp in length
            *     5816     5915: gap of 100 bp
            *     5916     6648: contig of 733 bp in length
            *     6649     6748: gap of 100 bp
            *     6749     7520: contig of 772 bp in length
            *     7521     7620: gap of 100 bp
            *     7621     8347: contig of 727 bp in length
            *     8348     8447: gap of 100 bp
            *     8448     9227: contig of 780 bp in length
            *     9228     9327: gap of 100 bp
            *     9328    10052: contig of 725 bp in length
            *    10053    10152: gap of 100 bp
            *    10153    10869: contig of 717 bp in length
            *    10870    10969: gap of 100 bp
            *    10970    11718: contig of 749 bp in length
            *    11719    11818: gap of 100 bp
            *    11819    12557: contig of 739 bp in length
            *    12558    12657: gap of 100 bp
            *    12658    13412: contig of 755 bp in length
            *    13413    13512: gap of 100 bp
            *    13513    14255: contig of 743 bp in length
            *    14256    14355: gap of 100 bp
            *    14356    15086: contig of 731 bp in length
            *    15087    15186: gap of 100 bp
            *    15187    15901: contig of 715 bp in length
            *    15902    16001: gap of 100 bp
            *    16002    16711: contig of 710 bp in length
            *    16712    16811: gap of 100 bp
            *    16812    17539: contig of 728 bp in length
            *    17540    17639: gap of 100 bp
            *    17640    18351: contig of 712 bp in length
            *    18352    18451: gap of 100 bp
            *    18452    19162: contig of 711 bp in length
            *    19163    19262: gap of 100 bp
            *    19263    20054: contig of 792 bp in length
            *    20055    20154: gap of 100 bp
            *    20155    20891: contig of 737 bp in length
            *    20892    20991: gap of 100 bp
            *    20992    21720: contig of 729 bp in length
            *    21721    21820: gap of 100 bp
            *    21821    22553: contig of 733 bp in length
            *    22554    22653: gap of 100 bp
            *    22654    23420: contig of 767 bp in length
            *    23421    23520: gap of 100 bp
            *    23521    24234: contig of 714 bp in length
            *    24235    24334: gap of 100 bp
            *    24335    25069: contig of 735 bp in length
            *    25070    25169: gap of 100 bp
            *    25170    25915: contig of 746 bp in length
            *    25916    26015: gap of 100 bp
            *    26016    26751: contig of 736 bp in length
            *    26752    26851: gap of 100 bp
            *    26852    27574: contig of 723 bp in length
            *    27575    27674: gap of 100 bp
            *    27675    28437: contig of 763 bp in length
            *    28438    28537: gap of 100 bp
            *    28538    29258: contig of 721 bp in length
            *    29259    29358: gap of 100 bp
            *    29359    30068: contig of 710 bp in length
            *    30069    30168: gap of 100 bp
            *    30169    30918: contig of 750 bp in length
            *    30919    31018: gap of 100 bp
            *    31019    31770: contig of 752 bp in length
            *    31771    31870: gap of 100 bp
            *    31871    32658: contig of 788 bp in length
            *    32659    32758: gap of 100 bp
```

```
            *    32759    33501: contig of 743 bp in length
            *    33502    33601: gap of 100 bp
            *    33602    34347: contig of 746 bp in length
            *    34348    34447: gap of 100 bp
            *    34448    35165: contig of 718 bp in length
            *    35166    35265: gap of 100 bp
            *    35266    35998: contig of 733 bp in length
            *    35999    36098: gap of 100 bp
            *    36099    36802: contig of 704 bp in length
            *    36803    36902: gap of 100 bp
            *    36903    37632: contig of 730 bp in length
            *    37633    37732: gap of 100 bp
            *    37733    38454: contig of 722 bp in length
            *    38455    38554: gap of 100 bp
            *    38555    39275: contig of 721 bp in length
            *    39276    39375: gap of 100 bp
            *    39376    40113: contig of 738 bp in length
            *    40114    40213: gap of 100 bp
            *    40214    40948: contig of 735 bp in length
            *    40949    41048: gap of 100 bp
            *    41049    41792: contig of 744 bp in length
            *    41793    41892: gap of 100 bp
            *    41893    42630: contig of 738 bp in length
            *    42631    42730: gap of 100 bp
            *    42731    43459: contig of 729 bp in length
            *    43460    43559: gap of 100 bp
            *    43560    44277: contig of 718 bp in length
            *    44278    44377: gap of 100 bp
            *    44378    45120: contig of 743 bp in length
            *    45121    45220: gap of 100 bp
            *    45221    45942: contig of 722 bp in length
            *    45943    46042: gap of 100 bp
            *    46043    46775: contig of 733 bp in length
            *    46776    46875: gap of 100 bp
            *    46876    47609: contig of 734 bp in length
            *    47610    47709: gap of 100 bp
            *    47710    48446: contig of 737 bp in length
            *    48447    48546: gap of 100 bp
            *    48547    49285: contig of 739 bp in length
            *    49286    49385: gap of 100 bp
            *    49386    50178: contig of 793 bp in length
            *    50179    50278: gap of 100 bp
            *    50279    51008: contig of 730 bp in length
            *    51009    51108: gap of 100 bp
            *    51109    51846: contig of 738 bp in length
            *    51847    51946: gap of 100 bp
            *    51947    52672: contig of 726 bp in length
            *    52673    52772: gap of 100 bp
            *    52773    53505: contig of 733 bp in length
            *    53506    53605: gap of 100 bp
            *    53606    54330: contig of 725 bp in length
            *    54331    54430: gap of 100 bp
            *    54431    55166: contig of 736 bp in length
            *    55167    55266: gap of 100 bp
            *    55267    55979: contig of 713 bp in length
            *    55980    56079: gap of 100 bp
            *    56080    56823: contig of 744 bp in length
            *    56824    56923: gap of 100 bp
            *    56924    57626: contig of 703 bp in length

  Query Match             33.5%;  Score 670.4;  DB 14;  Length 71032;
  Best Local Similarity   87.0%;  Pred. No. 5.7e-142;
  Matches  693;  Conservative    0;  Mismatches  102;  Indels    2;  Gaps    2;

Qy        131 CTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGGAAAGCCAAAA 190
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db      41937 CTTGTTTCCTTCCCTCCTC-AGGTTACCCACAATTCTCAGTTCTCTCAGGAAAGCCAAAA 41995

Qy        191 AATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTGATAATATGTT 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41996 AATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTGATAATATGTT 42055

Qy        251 CCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTGCATATGCAGG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42056 CCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTGCATATGCAGG 42115

Qy        311 AGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCCTGAACTGAGT 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42116 AGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCCTGAACTGAGT 42175

Qy        371 GAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATGTGCCAGTGGT 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42176 GAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATGTGCCAGTGGT 42235

Qy        431 CAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCACTGCTTCATT 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42236 CAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCACTGCTTCATT 42295

Qy        491 TCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTTCCCTCTGGCA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42296 TCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTTCCCTCTGGCA 42355

Qy        551 GTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGTACTCAAAGGA 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42356 GTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGTACTCAAAGGA 42415

Qy        611 GCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTATGGAAGCCAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42416 GCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTATGGAAGCCAA 42475

Qy        671 GGTTGCCATAGTGCCTCTT-GAGGTTGTTTGCTCAGCCAAGGCCCAAGCTTTGTGCTTCA 729
              |||||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db      42476 GGTTGCCATAGTGCCTCTTNGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTNTGTGCTTCA 42535

Qy        730 AACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCACCCAACTGG 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42536 AACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCACCCAACTGG 42595

Qy        790 ATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTTCAGGTGGGATGGTAGATGCTG 849
              ||||||||||||||||||||||||||||||||||
Db      42596 ATAAAAGAACAATTGCCATATCTCAATGACCACCTNNNNNNNNNNNNNNNNNNNNNNN 42655

Qy        850 GAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGCTCTGACTGGG 909

Db      42656 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 42715

Qy        910 GACCCTAAATATGCAAA 926
                            |
Db      42716 NNNNNNNNNNNNNNGA 42732
```

```
RESULT 10
AP006290/c
LOCUS       AP006290               27423 bp    DNA     linear   PRI 04-APR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 8 clone:RP11-1345M15, complete
            sequence.
ACCESSION   AP006290
VERSION     AP006290.2  GI:29539558
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2003)
REFERENCE   2  (bases 1 to 27423)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
```

```
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Apr 3, 2003 this sequence version replaced gi:29243350.
FEATURES             Location/Qualifiers
     source          1. .27423
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-1345M15"
ORIGIN

  Query Match             10.9%;  Score 218;  DB 8;  Length 27423;
  Best Local Similarity   68.4%;  Pred. No. 4.4e-39;
  Matches  361;  Conservative    0;  Mismatches  160;  Indels    7;  Gaps    4;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| |||||  |||||||||| ||||| ||||||| || ||||||||||||||| ||
Db       3905 TTTTTTTTTTTTTTCTTTTTTTGAGACAGATTCTCACTGTGACACCCAGGCTGGAGTGCA 3846

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             ||| |   | ||||||||||||||||||||||||||||||||||| | ||||||||
Db       3845 GTGAC-ACGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGTGCTTCTCCTGCC 3787

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTA 1391
             ||||| |||||| |||||||||||||||| | |||  |  |||  |  ||||||||||||
Db       3786 TCAGCCTCCCGAGTAGCTGGGATTACAGGCACCTGCCACCACATCCAGCTAATTTTTGTA 3727

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             |||||||||||||||| ||| | |||||||| ||||||| | |||||||||||||| |||
Db       3726 TTTTTAGTAGAGACAGGGTTCCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCA 3667

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
              ||||||||||  ||||||||||||||||  ||||||||||||||||||||||||| | |||
Db       3666 AGTGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGAGCC 3607

Qy       1512 CAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTC 1571
             | ||||   |       |     || |||||   ||        |  ||     | ||
Db       3606 CGGCCTTCCTGGGGTTCTTGATGATATGTTTTATAATGTCAGCCTCATACAAAATAATAG 3547

Qy       1572 ATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGGAAAATTCC 1631
              |||      |  |||   | | ||||  ||||  | ||||  |    ||| | |||   |
Db       3546 TTTTACAGCTTTTTACTATTTAATAAGCATTTTTTCCATCATAACATTTTAGCAAAA--C 3489

Qy       1632 CTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTTATGTTTCTTTC 1691
              || ||   |  |   | ||  |||  |   |  | | || ||  |  | | || ||
Db       3488 TTGCTTTTCTAATGGATGCTTAAAATCTCATTATATTTTAACTATTCCTCTATTGCTGGA 3429

Qy       1692 AAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTA 1739
              | ||||    || || || |    | |||| |    | |  || | |
Db       3428 TATTTATGTTGTTTTCAATTTGTACAATTATATATAACATTTTTATCA 3381


RESULT 11
AC136997/c
LOCUS       AC136997              133315 bp    DNA     linear   HTG 14-DEC-2002
DEFINITION  Homo sapiens chromosome 8 clone RP11-933I10 map 8, WORKING DRAFT
            SEQUENCE, 24 unordered pieces.
ACCESSION   AC136997
VERSION     AC136997.2  GI:26801292
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 133315)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-933I10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133315)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 133315)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 14, 2002 this sequence version replaced gi:24962720.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28557
                Center clone name: 933_I_10
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 125101 bases at least Q40
                Consensus quality: 128385 bases at least Q30
                Consensus quality: 129654 bases at least Q20
                Insert size: 206000; agarose-fp
                Insert size: 131015; sum-of-contigs
                Quality coverage: 15.8 in Q20 bases; agarose-fp
                Quality coverage: 24.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1      654: contig of 654 bp in length
            *      655      754: gap of 100 bp
            *      755     1683: contig of 929 bp in length
            *     1684     1783: gap of 100 bp
            *     1784     2094: contig of 311 bp in length
            *     2095     2194: gap of 100 bp
            *     2195     3125: contig of 931 bp in length
            *     3126     3225: gap of 100 bp
            *     3226     4290: contig of 1065 bp in length
            *     4291     4390: gap of 100 bp
            *     4391     5051: contig of 661 bp in length
            *     5052     5151: gap of 100 bp
            *     5152     7256: contig of 2105 bp in length
            *     7257     7356: gap of 100 bp
            *     7357     8674: contig of 1318 bp in length
            *     8675     8774: gap of 100 bp
            *     8775    10666: contig of 1892 bp in length
            *    10667    10766: gap of 100 bp
            *    10767    12819: contig of 2053 bp in length
            *    12820    12919: gap of 100 bp
            *    12920    14779: contig of 1860 bp in length
            *    14780    14879: gap of 100 bp
            *    14880    17960: contig of 3081 bp in length
            *    17961    18060: gap of 100 bp
            *    18061    21424: contig of 3364 bp in length
            *    21425    21524: gap of 100 bp
            *    21525    23085: contig of 1561 bp in length
            *    23086    23185: gap of 100 bp
            *    23186    28363: contig of 5178 bp in length
            *    28364    28463: gap of 100 bp
            *    28464    31735: contig of 3272 bp in length
            *    31736    31835: gap of 100 bp
            *    31836    39013: contig of 7178 bp in length
            *    39014    39113: gap of 100 bp
            *    39114    45836: contig of 6723 bp in length
            *    45837    45936: gap of 100 bp
            *    45937    63312: contig of 17376 bp in length
            *    63313    63412: gap of 100 bp
            *    63413    71785: contig of 8373 bp in length
            *    71786    71885: gap of 100 bp
            *    71886    82043: contig of 10158 bp in length
            *    82044    82143: gap of 100 bp
            *    82144    92091: contig of 9948 bp in length
            *    92092    92191: gap of 100 bp
            *    92192   106557: contig of 14366 bp in length
            *   106558   106657: gap of 100 bp
            *   106658   133315: contig of 26658 bp in length.
FEATURES             Location/Qualifiers
     source          1. .133315
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-933I10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .654
                     /note="assembly_fragment"
     gap             655. .754
                     /estimated_length=100
     misc_feature    755. .1683
                     /note="assembly_fragment"
     gap             1684. .1783
                     /estimated_length=100
     misc_feature    1784. .2094
                     /note="assembly_fragment"
     gap             2095. .2194
                     /estimated_length=100
     misc_feature    2195. .3125
                     /note="assembly_fragment"
     gap             3126. .3225
                     /estimated_length=100
     misc_feature    3226. .4290
                     /note="assembly_fragment"
     gap             4291. .4390
                     /estimated_length=100
     misc_feature    4391. .5051
                     /note="assembly_fragment"
     gap             5052. .5151
                     /estimated_length=100
     misc_feature    5152. .7256
                     /note="assembly_fragment"
     gap             7257. .7356
                     /estimated_length=100
     misc_feature    7357. .8674
                     /note="assembly_fragment"
     gap             8675. .8774
                     /estimated_length=100
     misc_feature    8775. .10666
                     /note="assembly_fragment"
     gap             10667. .10766
                     /estimated_length=100
     misc_feature    10767. .12819
                     /note="assembly_fragment"
     gap             12820. .12919
                     /estimated_length=100
     misc_feature    12920. .14779
                     /note="assembly_fragment"
     gap             14780. .14879
                     /estimated_length=100
     misc_feature    14880. .17960
                     /note="assembly_fragment"
     gap             17961. .18060
                     /estimated_length=100
     misc_feature    18061. .21424
                     /note="assembly_fragment"
     gap             21425. .21524
                     /estimated_length=100
     misc_feature    21525. .23085

  Query Match            10.9%;  Score 218;  DB 14;  Length 133315;
  Best Local Similarity  68.4%;  Pred. No. 3.6e-39;
  Matches  361;  Conservative    0;  Mismatches  160;  Indels    7;  Gaps    4;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| |||||||||| ||||| ||||||| || ||||||||||||||| ||
Db      41158 TTTTTTTTTTTTTTCTTTTTTTGAGACAGATTCTCACTGTGACACCCAGGCTGGAGTGCA 41099

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             ||| |   | ||||||||||||||||||||||||||||||||||||| | ||||||||
Db      41098 GTGAC-ACGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGTGCTTCTCCTGCC 41040

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTA 1391
             ||||| |||||| |||||||||||||||| | |||  |  ||| |  ||||||||||||
Db      41039 TCAGCCTCCCGAGTAGCTGGGATTACAGGCACCTGCCACCACATCCAGCTAATTTTTGTA 40980

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             |||||||||||||||| ||| | |||||||| ||||||| | |||||||||||||| |||
Db      40979 TTTTTAGTAGAGACAGGGTTCCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCA 40920

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
             ||||||||||  ||||||||||||||||  ||||||||||||||||||||||| | |||
Db      40919 AGTGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGAGCC 40860

Qy       1512 CAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTC 1571
             | ||||  |     |    || |||||  ||        |  ||    | ||
Db      40859 CGGCCTTCCTGGGGTTCTTGATGATATGTTTTATAATGTCAGCCTCATACAAAATAATAG 40800

Qy       1572 ATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGGAAAATTCC 1631
              |||       |  |||   | | ||||  ||||  | |||| |    ||| | |||   |
Db      40799 TTTTACAGCTTTTTACTATTTAATAAGCATTTTTTCCATCATAACATTTTAGCAAAA--C 40742

Qy       1632 CTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTTATGTTTCTTTC 1691
             || ||   |  |   | ||  |||  |   |  | | || ||  |  | | || ||
Db      40741 TTGCTTTTCTAATGGATGCTTAAAATCTCATTATATTTTAACTATTCCTCTATTGCTGGA 40682
```

```
Qy       1692 AAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTA 1739
              | ||||    || || || |    | |||| |    | |  || | |
Db      40681 TATTTATGTTGTTTTCAATTTGTACAATTATATATAACATTTTTATCA 40634


RESULT 12
AC130304/c
LOCUS       AC130304               161953 bp    DNA     linear   PRI 22-FEB-2003
DEFINITION  Homo sapiens chromosome 8, clone RP11-205M5, complete sequence.
ACCESSION   AC130304
VERSION     AC130304.12  GI:28467201
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 161953)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-205M5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161953)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 161953)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 161953)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 22, 2003 this sequence version replaced gi:27476171.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L27896
                Center clone name: 205_M_5
            --------------.
FEATURES             Location/Qualifiers
     source          1. .161953
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-205M5"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   830. .871
                     /rpt_family="(TA)n"
     repeat_region   complement(872. .1160)
                     /rpt_family="AluJb"
     repeat_region   complement(1901. .2187)
                     /rpt_family="AluSx"
     repeat_region   complement(2231. .2527)
                     /rpt_family="AluSx"
     unsure          3396. .3484
                     /note="PCR product sequence only"
     repeat_region   3689. .3841
                     /rpt_family="MIR3"
     repeat_region   3842. .4168
                     /rpt_family="AluJo"
     repeat_region   complement(4712. .4804)
                     /rpt_family="L2"
     repeat_region   5428. .5744
                     /rpt_family="AluSq"
     repeat_region   6094. .6127
                     /rpt_family="(GA)n"
     repeat_region   complement(6154. .6439)
                     /rpt_family="L2"
     repeat_region   complement(6553. .6658)
                     /rpt_family="L2"
     repeat_region   complement(6718. .6805)
                     /rpt_family="MER21B"
     repeat_region   complement(6806. .7107)
                     /rpt_family="AluSx"
     repeat_region   complement(7108. .7654)
                     /rpt_family="MER21B"
     repeat_region   complement(7789. .7814)
                     /rpt_family="MER21B"
     repeat_region   complement(7815. .8117)
                     /rpt_family="AluSq"
     repeat_region   complement(8118. .8188)
```

```
                     /rpt_family="MER21B"
     repeat_region   8255. .8781
                     /rpt_family="MER77"
     repeat_region   9010. .9367
                     /rpt_family="THE1B"
     repeat_region   complement(9952. .10253)
                     /rpt_family="AluSx"
     repeat_region   complement(10313. .10508)
                     /rpt_family="AluSx"
     repeat_region   complement(10545. .10856)
                     /rpt_family="L1MC5"
     repeat_region   10861. .10899
                     /rpt_family="AT_rich"
     repeat_region   10900. .11039
                     /rpt_family="FLAM_A"
     repeat_region   11070. .11157
                     /rpt_family="L1MD3"
     repeat_region   complement(11176. .11401)
                     /rpt_family="L1MC5"
     repeat_region   complement(11425. .11574)
                     /rpt_family="FRAM"
     repeat_region   complement(11589. .13219)
                     /rpt_family="SVA"
     repeat_region   13220. .13265
                     /rpt_family="(GGGAGA)n"
     repeat_region   complement(13412. .13711)
                     /rpt_family="AluJo"
     repeat_region   13761. .13896
                     /rpt_family="FLAM_C"
     repeat_region   13898. .13998
                     /rpt_family="GA-rich"
     repeat_region   14000. .14047
                     /rpt_family="(CGAGA)n"
     repeat_region   14067. .14192
                     /rpt_family="AluJb"
     repeat_region   14346. .14370
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(14758. .14968)
                     /rpt_family="L2"
     repeat_region   14972. .15220
                     /rpt_family="L1M4"
     repeat_region   15221. .15505
                     /rpt_family="AluSx"
     repeat_region   15506. .16451
                     /rpt_family="L1M4"
     repeat_region   16452. .16683
                     /rpt_family="AluSx"
     repeat_region   16684. .17223
                     /rpt_family="L1M4"
     repeat_region   complement(17224. .17557)
                     /rpt_family="MLT1A"
     repeat_region   17558. .17664
                     /rpt_family="L1M4"
     repeat_region   17665. .17989
                     /rpt_family="AluJo"
     repeat_region   17990. .18040
                     /rpt_family="L1M4"
     repeat_region   18041. .18356
                     /rpt_family="AluSq"
     repeat_region   18357. .18447
                     /rpt_family="L1M4"

  Query Match             10.9%;  Score 218;  DB 8;  Length 161953;
  Best Local Similarity   68.4%;  Pred. No. 3.5e-39;
  Matches  361;  Conservative    0;  Mismatches  160;  Indels    7;  Gaps    4;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| |||||  ||||||||| ||||| ||||||| || ||||||||||||||| ||
Db      74986 TTTTTTTTTTTTTTCTTTTTTTGAGACAGATTCTCACTGTGACACCCAGGCTGGAGTGCA 74927

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             ||| |   | ||||||||||||||||||||||||||||||||||||| | ||||||||
Db      74926 GTGAC-ACGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGTGCTTCTCCTGCC 74868

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTA 1391
             ||||| |||||| |||||||||||||||| | |||   |  ||| |  ||||||||||||
Db      74867 TCAGCCTCCCGAGTAGCTGGGATTACAGGCACCTGCCACCACATCCAGCTAATTTTTGTA 74808

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             ||||||||||||||| ||| | |||||||| ||||||| | ||||||||||||| |||
Db      74807 TTTTTAGTAGAGACAGGGTTCCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCA 74748

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
             |||||||||| |||||||||||||||  ||||||||||||||||||||||| | |||
Db      74747 AGTGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGAGCC 74688

Qy       1512 CAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTC 1571
             | ||||  |       |     || |||||  ||        |  ||     | ||
Db      74687 CGGCCTTCCTGGGGTTCTTGATGATATGTTTTATAATGTCAGCCTCATACAAAATAATAG 74628

Qy       1572 ATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGGAAAATTCC 1631
             |||      |  |||   | | ||||  ||||  | |||| |   ||| | |||   |
Db      74627 TTTTACAGCTTTTTACTATTTAATAAGCATTTTTTCCATCATAACATTTTAGCAAAA--C 74570

Qy       1632 CTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTTATGTTTCTTTC 1691
             || ||   |  |   | ||  |||  |   |  | | || ||  |  | | || ||
Db      74569 TTGCTTTTCTAATGGATGCTTAAAATCTCATTATATTTTAACTATTCCTCTATTGCTGGA 74510

Qy       1692 AAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTA 1739
             | ||||   || || || |    | |||| |    | |  || | |
Db      74509 TATTTATGTTGTTTTCAATTTGTACAATTATATATAACATTTTTATCA 74462


RESULT 13
AL353716/c
LOCUS       AL353716               98340 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-501I18 on chromosome 6 Contains
            the TBCC gene for tubulin-specific chaperone c, part of a novel
            gene, part of the gene for a novel protein (KIAA0240) and CpG
            islands, complete sequence.
ACCESSION   AL353716
VERSION     AL353716.18  GI:11967504
KEYWORDS    HTG; KIAA0240; TBCC; tubulin.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 98340)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 22, 2000 this sequence version replaced gi:11611339.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-501I18 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
```

```
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one subclone; and the assembly was confirmed by restriction digest,
          except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .98340
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-501I18"
                     /clone_lib="RPCI-11.2"
     misc_feature    100
                     /note="Clone_right_end: RP3-392M17"
     gene            complement(4427. .6042)
                     /gene="TBCC"
                     /locus_tag="RP11-501I18.1-001"
     mRNA            complement(4427. .6042)
                     /gene="TBCC"
                     /locus_tag="RP11-501I18.1-001"
                     /note="match: cDNAs: AK012950.1 U61234.1"
     polyA_site      complement(4427)
                     /gene="TBCC"
                     /locus_tag="RP11-501I18.1-001"
     gene            complement(join(4444. .6041,6226. .6766))
                     /gene="TBCC"
                     /locus_tag="RP11-501I18.1-002"
     mRNA            complement(join(4444. .6041,6226. .6766))
                     /gene="TBCC"
                     /locus_tag="RP11-501I18.1-002"
                     /note="match: cDNAs: BC017479"
     polyA_signal    complement(4460. .4465)
                     /gene="TBCC"
                     /locus_tag="RP11-501I18.1-001"
     CDS             complement(4979. .6019)
                     /gene="TBCC"
                     /locus_tag="RP11-501I18.1-001"
                     /standard_name="OTTHUMP00000016406"
                     /note="match: proteins: Q15814 Q9SMR2"
                     /codon_start=1
                     /protein_id="CAI14818.1"
                     /db_xref="GI:55958597"
                     /db_xref="GOA:Q5T787"
                     /db_xref="InterPro:IPR006599"
                     /db_xref="UniProt/TrEMBL:Q5T787"
                     /translation="MESVSCSAAAVRTGDMESQRDLSLVPERLQRREQERQLEVERRK
                     QKRQNQEVEKENSHFFVATFVRERAAVEELLERAESVERLEEAASRLQGLQKLINDSV
                     FFLAAYDLRQGQEALARLQAALAERRRGLQPKKRFAFKTRGKDAASSTKVDAAPGIPP
                     AVESIQDSPLPKKAEGDLGPSWVCGFSNLESQVLEKRASELHQRDVLLTELSNCTVRL
                     YGNPNTLRLTKAHSCKLLCGPVSTSVFLEDCSDCVLAVACQQLRIHSTKDTRIFLQVT
                     SRAIVEDCSGIQFAPYTWSYPEIDKDFESSGLDRSKNNWNDVDDFNWLARDMASPNWS
                     ILPEEERNIQWD"
     misc_feature    41953. .41969
                     /note="Weak data. Sequence from dGTP reads only."
     gene            join(49206. .49757,50088. .51265)
                     /locus_tag="RP11-501I18.2-001"
     mRNA            join(49206. .49757,50088. .51265)
                     /locus_tag="RP11-501I18.2-001"
                     /note="match: cDNAs: AK024196.1"
     gene            join(<88566. .90118,complement(AL035587.5:82845. .99312))
                     /gene="KIAA0240"
                     /locus_tag="RP3-475N16.5-001"
     mRNA            join(<88566. .90118,complement(AL035587.5:99207. .99312),
                     complement(AL035587.5:97666. .97766),
                     complement(AL035587.5:95487. .95548),
                     complement(AL035587.5:94168. .94313),
                     complement(AL035587.5:91048. .91142),
                     complement(AL035587.5:88792. .88894),
                     complement(AL035587.5:82845. .86745))
                     /gene="KIAA0240"
                     /locus_tag="RP3-475N16.5-001"
                     /note="match: ESTs: AA027241 AA057261 AA058706 AA265684
                     AA318197 AA429191 AA493882 AA551508 AA551530 AA766531
                     AA822777 AI032192 AI069983 AI256570 AI303778 AI430383
                     AI466036 AI493016 AI628139 AI693499 AI717510 AI801951
                     AI806974 AI911090 AL048597 AU059481 AW008530 AW016691
                     AW026001 AW028017 H14203 H14204 H27629 N51260 W93617
                     match: cDNAs: D87077 D87077.1"
     CDS             join(<88566. .90118,complement(AL035587.5:99207. .99312),
                     complement(AL035587.5:97666. .97766),
                     complement(AL035587.5:95487. .95548),
                     complement(AL035587.5:94168. .94313),
                     complement(AL035587.5:91048. .91142),
                     complement(AL035587.5:88792. .88894),
                     complement(AL035587.5:85958. .86745))
                     /gene="KIAA0240"
                     /locus_tag="RP3-475N16.5-001"
                     /standard_name="OTTHUMP00000016407"
                     /note="match: proteins: Q92514"
                     /codon_start=3
                     /protein_id="CAI14820.1"
                     /db_xref="GI:55958599"
                     /db_xref="Genew:21111"
                     /db_xref="UniProt/TrEMBL:Q92514"
                     /translation="LTEDQPFDILQKSLQEANITEQTLAEEAYLDASIGSSQQFAQAQ
                     LHPSSSASFTQASNVSNYSGQTLQPIGVTHVPVGASFASNTVGVQHGFMQHVGISVPS
                     QHLSNSSQISGSGQIQLIGSFGNHPSMMTINNLDGSQIILKGSGQQAPSNVSGGLLVH
                     RQTPNGNSLFGNSSSSPVAQPVTVPFNSTNFQTSLPVHNIIIQRGLAPNSNKVPINIQ
                     PKPIQMGQQNTYNVNNLGIQQHHVQQGISFASASSPQGSVVGPHMSVNIVNQQNTRKP
                     VTSQAVSSTGGSIVIHSPMGQPHAPQSQFLIPTSLSVSSNSVHHVQTINGQLLQTQPS
                     QLISGQVASEHVMLNRNSSNMLRTNQPYTGPMLNNQNTAVHLVSGQTFAASGSPVIAN
                     HASPQLVGGQMPLQQASPTVLHLSPGQSSVSQGRPGFATMPSVTSMSGPSRFPAVSSA
                     STAHPSLGSAVQSGSSGSNFTGDQLTQPNRTPVPVSVSHRLPVSSSKSTSTFSNTPGT
                     GTQQQFFCQAQKKCLNQTSPISAPKTTDGLRQAQIPGLLSTTLPGQDSGSKVISASLG
                     TAQPQQEKVVGSSPGHPAVQVESHSGGQKRPAAKQLTKGAFILQQLQRDQAHTVTPDK
                     SHFRSLSDAVQRLLSYHVCQGSMPTEEDLRKVDNEFETVATQLLKRTQAMLNKYRCLL
                     LEDAMRINPSAEMVMIDRMFNQEERASLSRDKRLALVDPEGFQADFCCSFKLDKAAHE
                     TQFGRSDQHGSKASSSLQPPAKAQGRDRAKTGVTEPMNHDQFHLVPNHIVVSAEGNIS
                     KKTECLGRALKFDKVGLVQYQSTSEEKASRREPLKASQCSPGPEGHRKTSSRSDHGTE
                     SKLSSILADSHLEMTCNNSFQDKSLRNSPKNEVLHTDIMKGSGEPQPDLQLTKSLETT
                     FKNILELKKAGRQPQSDPTVSGSVELDFPNFSPMASQENCLEKFIPDHSEGVVETDSI
                     LEAAVNSILEC"
     misc_feature    98241
                     /note="Clone_left_end: RP3-475N16"
ORIGIN

  Query Match           10.8%;  Score 216.6;  DB 8;  Length 98340;
  Best Local Similarity   79.8%;  Pred. No. 7.7e-39;
  Matches  293;  Conservative    0;  Mismatches   69;  Indels    5;  Gaps    3;

Qy       1212 CAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAG 1271
             |||| ||||| ||||| |||||| |||||| ||||| || |||| ||||||||||||
Db      24213 CAGACATTTTTGTTTTTTTTTTTTCGGAAACAAAGTCTTGCTCTGTCGCCCAGGCTGGAG 24154

Qy       1272 TACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCC 1331
             ||||||||| || ||| ||||||||||||||||||||||||||||||||| ||||||||
Db      24153 TACAGTGGC-ATGATCTTGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCC 24095

Qy       1332 TG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTT 1387
             ||  ||||| |||||| ||||||||| |||||||| |||  |  ||   | ||||||||
Db      24094 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCGTGCCACCACACCCGGCTAATTTT 24035

Qy       1388 TGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGAC 1447
             |||||||||||||||||| ||||| |||||||| ||||||| | ||||||||||||||
Db      24034 TGTATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGAC 23975

Qy       1448 TTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCG 1507
             |||||||||| ||| |||| |||||||||||  |||||||||||||  |||||||| |
Db      23974 CTCAGGTGATCTGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTG 23915

Qy       1508 TGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTT 1567
             ||||||||||       ||  |  | | |  |||||   | |||  ||| | | || |
Db      23914 TGCCCAGCCTTGGCAGACATTTAATATGCTATTTTCTTTTGTCTTTTTTTTTTTTTTTA 23855

Qy       1568 ATTCATT 1574
```

```
              || || |
Db      23854 ATGCACT 23848


RESULT 14
AL353694_3
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AL353694 Accession AL353694
     Fragment Name      Begin     End
     AL353694_0             1    110000
     AL353694_1        100001    210000
     AL353694_2        200001    310000
     AL353694_3        300001    356330
Continuation (4 of 4) of AL353694 from base 300001 (AL353694 Homo sapiens chromosome 6 c

  Query Match            10.8%;  Score 215.8;  DB 14;  Length 56330;
  Best Local Similarity  82.1%;  Pred. No. 1.3e-38;
  Matches  285;  Conservative    0;  Mismatches   57;  Indels    5;  Gaps    3;

Qy       1192 AGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCA 1251
              || | |  |||    | ||||   | ||||| ||||| |||||||||| || |||||||
Db      21178 AGAAGTGCATTCATCTCTCTCTTTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCG 21237

Qy       1252 CTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTC 1311
              || |||||||||||||||||| |||||| || | ||||||||||||||||||||  ||||
Db      21238 CTCTGTCACCCAGGCTGGAGTGCAGTGG-TGCGATCTCGGCTCACTGCAACCTCCACCTC 21296

Qy       1312 CCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC 1369
              || ||||||||||||||||||  ||||| ||| || |||||||||||||||||| |||
Db      21297 CCGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCGCCTGC 21356

Qy       1370 ACCACCA--TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAG 1427
                |  ||  || |||||||||||||||||||||||||| | ||||| |||||||| ||||
Db      21357 CACCACACCTGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAG 21416

Qy       1428 GCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGG 1487
              ||| | || |||||||||||| |||| |||||||||||||  ||||||||||  |||||
Db      21417 GCTGGTCTCGAACTCCTGACCTCAGATGATCCACCCACCTTGGCCTCCCAAAGTGCTGGG 21476

Qy       1488 ATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGT 1534
              ||||||||| ||||||||| ||||||||||||    |||  || | |
Db      21477 ATTACAGGCGTGAGCCACCATGCCCAGCCTGTGCATTCATCTCTTTT 21523


RESULT 15
AC019047
LOCUS       AC019047               165007 bp    DNA     linear   PRI 15-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-83N2 from 2, complete sequence.
ACCESSION   AC019047
VERSION     AC019047.7  GI:16756327
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 165007)
  AUTHORS   Maupin,R., McLellan,M. and Griffin,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-83N2
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 165007)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 165007)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 165007)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 7, 2001 this sequence version replaced gi:13518242.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0083N02
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-485M18; the clone sequenced
            to the right is RP11-122C15.  Actual start of this clone is at base
            position 1 of RP11-83N2; actual end is at base position 165007 of
            RP11-83N2.

            The region from base position 29896 to 29936 is only covered by
            data from AC079400.
FEATURES             Location/Qualifiers
     source          1. .165007
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-83N2"
                     /clone_lib="RPCI-11"
     gene            67312. .155998
                     /gene="CENTG2"
     mRNA            join(67312. .67466,71502. .71592,79810. .80032,
                     150943. .151198,154670. .155998)
                     /gene="CENTG2"
     CDS             join(<67312. .67466,71502. .71592,79810. .80032,
                     150943. .151198,154670. .154873)
                     /gene="CENTG2"
                     /note="Homo sapiens centaurin, gamma 2 (CENTG2), mRNA.;
                     H_NH0083N02.1
                     This gene was based on gi(7662483)
                     Continued from H_NH0485M18.1"
                     /codon_start=3
                     /product="unknown"
                     /protein_id="AAX88852.1"
                     /db_xref="GI:62630106"
                     /translation="QEENFEFIIVSLTGQTWHFEATTYEERDAWVQAIESQILASLQS
```

```
                     CESSKNKSRLTSQSEAMALQSIRNMRGNSHCVDCETQNPNWASLNLGALMCIECSGIH
                     RNLGTHLSRVRSLDLDDWPVELIKVMSSIGNELANSVWEESSQGRTKPSVDSTREEKE
                     RWIRAKYEQKLFLAPLPCTELSLGQHLLRATADEDLRTAILLLAHGSRDEVNETCGEG
                     DGRTALHLACRKGNVVLAQLLIWYGVDVTARDAHGNTALAYARQASSQECIDVLLQYG
                     CPDERFVLMATPNLSRRNNNRNNSSGRVPTII"
     misc_feature    99273. .99544
                     /gene="CENTG2"
                     /note="CpG_island (%GC=61.8, o/e=0.82, #CpGs=23)"
     misc_feature    132326. .132557
                     /gene="CENTG2"
                     /note="CpG_island (%GC=57.3, o/e=0.95, #CpGs=20)"
     misc_feature    150956. .151228
                     /gene="CENTG2"
                     /note="CpG_island (%GC=70.3, o/e=0.71, #CpGs=25)"
     misc_feature    154665. .154963
                     /gene="CENTG2"
                     /note="CpG_island (%GC=65.9, o/e=0.83, #CpGs=29)"
ORIGIN

  Query Match             10.8%;  Score 215.2;  DB 8;  Length 165007;
  Best Local Similarity   75.9%;  Pred. No. 1.5e-38;
  Matches  305;  Conservative    0;  Mismatches   93;  Indels    4;  Gaps    3;

Qy       1126 ATAAGGAGATTTTTTAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTT 1185
             | | || |||| |  | |  || || |  |||           | | ||  |     |  |
Db     149047 AGAGGGCGATTGTGAATCAGTGCGTTATTAAAGTCTCTATGGGTAACTGCAGTCCCTTCT 149106

Qy       1186 CAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGA 1245
              | |       |  |  |  |||      |||||| ||||| ||||||| || |  ||
Db     149107 TACATGTAGGGCTTTTCTTTCTTTTCTTTTTCTTTTTTTTTTTTTTTTTTCGAGATGGA 149166

Qy       1246 GTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTC 1305
              |||||| | ||| |||||||||||||| ||||||| |   |||| ||||||||||||||
Db     149167 GTCTCATTCTGTTACCCAGGCTGGAGTGCAGTGGC-GCAATCTCAGCTCACTGCAACCTC 149225

Qy       1306 TGCCTCCCAGGTTCAAGCGATTCTCCT--GTCAGCTTCCCGAATAGCTGGGATTACAGGC 1363
              |||||||| ||||||||| ||||||||   || ||||||||| |||||||||||||||||
Db     149226 TGCCTCCCGGGTTCAAGCCATTCTCCTACCTCGGCTTCCCGAGTAGCTGGGATTACAGGC 149285

Qy       1364 GCATG-CACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTG 1422
              || || ||||||   | ||||||||||||||||||||||||||| ||||| ||||||||
Db     149286 GCCTGCCACCACACCGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTG 149345

Qy       1423 ACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCA 1482
               ||| ||| | || ||||||||||| |||||||||||||||||||| ||||||||||
Db     149346 GCCAAGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTG 149405

Qy       1483 CTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCT 1524
              |||||||||||||||||||||| | ||||||||| ||||||
Db     149406 CTGGGATTACAGGCATGAGCCACTGCGCCCAGCCTCTTTTCT 149447


RESULT 16
AP002889
LOCUS       AP002889              172296 bp    DNA     linear   HTG 18-OCT-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-148E9 map 11q, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AP002889
VERSION     AP002889.1  GI:10880445
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 172296)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 172,296 genomic DNA of 11q
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 172296)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     -------------- Genome Center
              Center: RIKEN Genomic Sciences Center(GSC)
              Center code: RIKEN
              Web site: http://hgp.gsc.riken.go.jp/
              Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
              Center project name: HumDraft11
              Center clone name: RP11-148E9
            -------------- Summary Statistics
              Sequencing vector: PCR products; 100% of reads
              Chemistry: Dye-terminator ET-amersham; 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 164750 bases at least Q40
              Consensus quality: 168344 bases at least Q30
              Consensus quality: 169784 bases at least Q20
              Insert size: 170496; sum-of-contigs
              Quality coverage: 5.98x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            19 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     37986 contig of  37986 bp in length
              38087     62025 contig of  23939 bp in length
              62126     73587 contig of  11462 bp in length
              73688     84480 contig of  10793 bp in length
              84581     96212 contig of  11632 bp in length
              96313    105504 contig of   9192 bp in length
             105605    112728 contig of   7124 bp in length
             112829    120174 contig of   7346 bp in length
             120275    129224 contig of   8950 bp in length
             129325    137506 contig of   8182 bp in length
             137607    144260 contig of   6654 bp in length
             144361    150760 contig of   6400 bp in length
             150861    154751 contig of   3891 bp in length
             154852    158790 contig of   3939 bp in length
             158891    163714 contig of   4824 bp in length
             163815    166989 contig of   3175 bp in length
             167090    169622 contig of   2533 bp in length
             169723    171188 contig of   1466 bp in length
             171289    172296 contig of   1008 bp in length.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    37986: contig of 37986 bp in length
            *   37987    38086: gap of 100 bp
            *   38087    62025: contig of 23939 bp in length
            *   62026    62125: gap of 100 bp
            *   62126    73587: contig of 11462 bp in length
            *   73588    73687: gap of 100 bp
            *   73688    84480: contig of 10793 bp in length
            *   84481    84580: gap of 100 bp
            *   84581    96212: contig of 11632 bp in length
            *   96213    96312: gap of 100 bp
            *   96313   105504: contig of 9192 bp in length
            *  105505   105604: gap of 100 bp
            *  105605   112728: contig of 7124 bp in length
            *  112729   112828: gap of 100 bp
            *  112829   120174: contig of 7346 bp in length
```

```
           *   120175   120274: gap of 100 bp
           *   120275   129224: contig of 8950 bp in length
           *   129225   129324: gap of 100 bp
           *   129325   137506: contig of 8182 bp in length
           *   137507   137606: gap of 100 bp
           *   137607   144260: contig of 6654 bp in length
           *   144261   144360: gap of 100 bp
           *   144361   150760: contig of 6400 bp in length
           *   150761   150860: gap of 100 bp
           *   150861   154751: contig of 3891 bp in length
           *   154752   154851: gap of 100 bp
           *   154852   158790: contig of 3939 bp in length
           *   158791   158890: gap of 100 bp
           *   158891   163714: contig of 4824 bp in length
           *   163715   163814: gap of 100 bp
           *   163815   166989: contig of 3175 bp in length
           *   166990   167089: gap of 100 bp
           *   167090   169622: contig of 2533 bp in length
           *   169623   169722: gap of 100 bp
           *   169723   171188: contig of 1466 bp in length
           *   171189   171288: gap of 100 bp
           *   171289   172296: contig of 1008 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172296
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-148E9"
     misc_feature    1. .37986
                     /note="assembly_fragment"
     misc_feature    38087. .62025
                     /note="assembly_fragment"
     misc_feature    62126. .73587
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    73688. .84480
                     /note="assembly_fragment"
     misc_feature    84581. .96212
                     /note="assembly_fragment"
     misc_feature    96313. .105504
                     /note="assembly_fragment"
     misc_feature    105605. .112728
                     /note="assembly_fragment"
     misc_feature    112829. .120174
                     /note="assembly_fragment"
     misc_feature    120275. .129224
                     /note="assembly_fragment"
     misc_feature    129325. .137506
                     /note="assembly_fragment"
     misc_feature    137607. .144260
                     /note="assembly_fragment"
     misc_feature    144361. .150760
                     /note="assembly_fragment"
     misc_feature    150861. .154751
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    154852. .158790
                     /note="assembly_fragment"
     misc_feature    158891. .163714
                     /note="assembly_fragment"
     misc_feature    163815. .166989
                     /note="assembly_fragment"
     misc_feature    167090. .169622
                     /note="assembly_fragment"
     misc_feature    169723. .171188
                     /note="assembly_fragment"
     misc_feature    171289. .172296
                     /note="assembly_fragment"
ORIGIN

  Query Match           10.7%;  Score 215;  DB 14;  Length 172296;
  Best Local Similarity   81.5%;  Pred. No. 1.7e-38;
  Matches  286;  Conservative    0;  Mismatches   60;  Indels    5;  Gaps    3;

Qy       1196 TTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTT 1255
             ||| ||   |||| ||    | ||||| ||||| |||||||||| || ||| ||| || |
Db      73689 TTTTTTTCTTTTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGCCTCGCTCT 73748

Qy       1256 GTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAG 1315
             ||| |||||||||||||| || ||||  || ||||||||||||||||||||||||||| |
Db      73749 GTCTCCCAGGCTGGAGTGCAATGGC-ATGATCTCGGCTCACTGCAACCTCTGCCTCCCGG 73807

Qy       1316 GTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCA 1373
             |||||||||||||||||  | ||| |||| | |||||||||| ||||  |||  |||||
Db      73808 GTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGTAGCTGGGATCACAG--GCACCCACCA 73865

Qy       1374 CCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGC 1433
             |  ||||| |||||||||||||||||||| | ||||| |||||||| ||||||| |
Db      73866 CTGTGCCTGATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGT 73925

Qy       1434 CTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACA 1493
              | ||||||||||  |||||||||||||||| |||||||||||||||  ||||||||||
Db      73926 TTCGAACTCCTGATCTCAGGTGATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACA 73985

Qy       1494 GGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCT 1544
             ||| |||||||||| |||| |||| |||   |    || | ||||| ||| |
Db      73986 GGCGTGAGCCACCGCGCCCGGCCTCTTTCTTTCTTTCTTTTATTTATTTTT 74036


RESULT 17
AP000221/c
LOCUS       AP000221               89323 bp    DNA     linear   PRI 20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 21q21.2, LL56-APP region,
            clone:S672, complete sequence.
ACCESSION   AP000221
VERSION     AP000221.1  GI:4835590
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 89323)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 89,323bp genomic DNA of 21q21.2
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 89323)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     The sequence is a part of the data (ACCESSION No. AP000136 -
            AP000145).
            The sequencing project is supported by Japan Science Technology
            Corporation (JST) and The Institute of Physical and Chemical
            Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .89323
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.2"
                     /clone="S672"
ORIGIN

  Query Match           10.7%;  Score 214;  DB 8;  Length 89323;
  Best Local Similarity   72.0%;  Pred. No. 3.1e-38;
  Matches  321;  Conservative    0;  Mismatches  120;  Indels    5;  Gaps    3;
```

```
Qy       1167 TTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTT 1226
Db      41407 TTGAATTGATAATTTTGTCCAGAGAAAACCTTGGATTTTTTTTTTTTTTTTTTTTTTTTT 41348

Qy       1227 TTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGT 1286
Db      41347 TTTTTTTTTTGAGAAGGAGTCTCGCTCTGTCACTCAGGCTGGAGTGCATTGG-TGCGAT 41289

Qy       1287 CTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCG 1344
Db      41288 CTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTTCTGCCTCAGCCTCCTG 41229

Qy       1345 AATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGA 1402
Db      41228 AGTAGCTGGGACTACAGGAGCGTGCCACCACTCCCAGCTAACTTTTGTATTTTTAGTAGA 41169

Qy       1403 GACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACC 1462
Db      41168 GAAAGGGTTTCACCATATTGGTCAGTCTGGTCTCGAACTCCTGACCTCAGGTGATGCACC 41109

Qy       1463 CACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTT 1522
Db      41108 CACCTCAGCCTCCCAAAGTGCTGGGACTACAGGCATGAGCCACCATGCCCGGCTGGATAT 41049

Qy       1523 CTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTCATTTTGGAAGT 1582
Db      41048 TTAACTTATTATATTTTATTGCTAGCTATTAATGCTTTCTATATGCTAAGCATTCAAAGT 40989

Qy       1583 AGTACACCTAAGTAAGGTTTTTAACA 1608
Db      40988 CATTTACAAATATTAACTGCTTAAAA 40963
```

```
RESULT 18
AP000085/c
LOCUS       AP000085              100000 bp    DNA     linear   PRI 25-MAY-2002
DEFINITION  Homo sapiens genomic DNA of 21q22.1, APP related, B2291C14-T1533
            region, segment 2/7.
ACCESSION   AP000085
VERSION     AP000085.1  GI:4730826
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens 782,611bp genomic DNA of 21q22.1 APP region
  JOURNAL   Published Only in Database (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Mika Hirakawa, Japan Science and Technology
            Corporation (JST), Advanced Databases Department; 5-3, Yonbancho,
            Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mika@tokyo.jst.go.jp,
            URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
            Fax:81-3-5214-8470)
COMMENT     This sequence is conducted by Kitasato University JST sequencing
            Laboratory as a JST sequencing team.
            Principal Investigator:Yoshiyuki Sakaki Ph.D.
            Phone: +81-3-5449-5622 , Fax: +81-3-5449-5445
            sakaki@hgc.ims.u-tokyo.ac.jp
            Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
            sequence is submitted by Human Genome Sequencing in ALIS project of
            JST.
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/) or send
            email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES              Location/Qualifiers
     source           1. .100000
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q22.1"
     STS              53504. .53759
                      /standard_name="D21S265"
                      /note="AFM234wa5;Genethon Marker;The location is between
                      each flanking site of PCR primers."
                      /db_xref="GDB:188463"
ORIGIN

  Query Match           10.7%;  Score 214;  DB 8;  Length 100000;
  Best Local Similarity   72.0%;  Pred. No. 3e-38;
  Matches  321;  Conservative    0;  Mismatches  120;  Indels    5;  Gaps    3;

Qy       1167 TTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTT 1226
Db       1909 TTGAATTGATAATTTTGTCCAGAGAAAACCTTGGATTTTTTTTTTTTTTTTTTTTTTTTT 1850

Qy       1227 TTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGT 1286
Db       1849 TTTTTTTTTTGAGAAGGAGTCTCGCTCTGTCACTCAGGCTGGAGTGCATTGG-TGCGAT 1791

Qy       1287 CTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCG 1344
Db       1790 CTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTTCTGCCTCAGCCTCCTG 1731

Qy       1345 AATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGA 1402
Db       1730 AGTAGCTGGGACTACAGGAGCGTGCCACCACTCCCAGCTAACTTTTGTATTTTTAGTAGA 1671

Qy       1403 GACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACC 1462
Db       1670 GAAAGGGTTTCACCATATTGGTCAGTCTGGTCTCGAACTCCTGACCTCAGGTGATGCACC 1611

Qy       1463 CACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTT 1522
Db       1610 CACCTCAGCCTCCCAAAGTGCTGGGACTACAGGCATGAGCCACCATGCCCGGCTGGATAT 1551

Qy       1523 CTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTCATTTTGGAAGT 1582
Db       1550 TTAACTTATTATATTTTATTGCTAGCTATTAATGCTTTCTATATGCTAAGCATTCAAAGT 1491

Qy       1583 AGTACACCTAAGTAAGGTTTTTAACA 1608
Db       1490 CATTTACAAATATTAACTGCTTAAAA 1465
```

```
RESULT 19
AP000137/c
LOCUS       AP000137              100000 bp    DNA     linear   PRI 08-JAN-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q21.2, LL56-APP region,
            clone B2291C14-R44F3, segment 2/10, complete sequence.
ACCESSION   AP000137
VERSION     AP000137.1  GI:4827103
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 911,949bp genomic DNA of 21q21.2 (REGION: LL56-APP
            CLONE RANGE: B2291C14-R44F3)
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (10-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     E. coli transposon insertion:The present data does not contain E.
            coli transposon sequences which integrated in the
            original/previous sequences.  We determined the boundary between
            the insertion and genomic sequences experimentally, removed the
            insertion sequences, reconstituted the present data. The sequencing
            project is supported by Japan Science Technology Corporation (JST)
            and The Institute of Physical and Chemical Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.2"
ORIGIN

  Query Match           10.7%;  Score 214;  DB 8;  Length 100000;
  Best Local Similarity   72.0%;  Pred. No. 3e-38;
  Matches  321;  Conservative    0;  Mismatches  120;  Indels    5;  Gaps    3;

Qy       1167 TTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTT 1226
             ||  | || | |  |||| |||||||    |||  | | |||| |    | ||||| |||
Db       1909 TTGAATTGATAATTTTGTCCAGAGAAAACCTTGGATTTTTTTTTTTTTTTTTTTTTTTTT 1850

Qy       1227 TTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGT 1286
             ||  |||||||||| |  ||||||| || |||||| |||||||||| || ||| || | |
Db       1849 TTTTTTTTTTTGAGAAGGAGTCTCGCTCTGTCACTCAGGCTGGAGTGCATTGG-TGCGAT 1791

Qy       1287 CTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCG 1344
             |||||||||||||||||||||||||  ||||||||||||||| |||  ||||| |||.|
Db       1790 CTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTTCTGCCTCAGCCTCCTG 1731

Qy       1345 AATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGA 1402
              | ||||||||| |||||| || |||   |  |    |  |||| ||||||||||||||||||
Db       1730 AGTAGCTGGGACTACAGGAGCGTGCCACCACTCCCAGCTAACTTTTGTATTTTTAGTAGA 1671

Qy       1403 GACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACC 1462
             || || ||||| |||| |||  ||| || | || |||||||||| |||||||| ||||
Db       1670 GAAAGGGTTTCACCATATTGGTCAGTCTGGTCTCGAACTCCTGACCTCAGGTGATGCACC 1611

Qy       1463 CACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTT 1522
             ||||||||||||||||  |||||| ||||||||||||||| |||||  ||  | | |
Db       1610 CACCTCAGCCTCCCAAAGTGCTGGGACTACAGGCATGAGCCACCATGCCCGGCTGGATAT 1551

Qy       1523 CTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTCATTTTGGAAGT 1582
              | |  | | |||||  ||   |    || ||   |    | | |     ||  ||||
Db       1550 TTAACTTATTATATTTTATTGCTAGCTATTAATGCTTTCTATATGCTAAGCATTCAAAGT 1491

Qy       1583 AGTACACCTAAGTAAGGTTTTTAACA 1608
               |  ||  |  | |  |  |||| |
Db       1490 CATTTACAAATATTAACTGCTTAAAA 1465


RESULT 20
AP001693/c
LOCUS       AP001693               340000 bp    DNA     linear   PRI 21-MAY-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, section 37/105.
ACCESSION   AP001693 AL163238 BA000005
VERSION     AP001693.1  GI:7768672
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,K., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     The DNA sequence of human chromosome 21
  JOURNAL   Nature 405 (6784), 311-319 (2000)
  PUBMED    10830953
REFERENCE   2  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
COMMENT     On May 30, 2000 this sequence version replaced gi:7717291.
            The chromosome 21 mapping and sequencing consortium consisting of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: hattori@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Molecular Biology, * Tokyo
            160-8582, Japan,
            * e.mail: nshimizu@dmb-med.keio.ac.jp
            * URL: http://www.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/
            AL163238: Submitted (10-Apr-2000).
FEATURES             Location/Qualifiers
     source          1. .340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
     source          <1. .63575
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
                     /chromosome="21"
                     /map="21q21.2"
                     /clone="R266G18, 5' partial"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AP001341"
     source          63192. .121962
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.2"
                     /clone="CTD2289H10"
                     /clone_lib="CTD BAC library"
                     /note="Accession No. AP001340"
     source          121433. .140060
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.2"
                     /clone="f39G8"
                     /clone_lib="CMF21 Fosmid library"
                     /note="Accession No. AP001348"
     source          139705. .200913
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.2"
                     /clone="CTD2291C14"
                     /clone_lib="CTD BAC library"
                     /note="Accession No. AP000220"
     source          200207. .289529
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.2"
                     /clone="pS672"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000221"
     source          271109. .311125
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.2"
                     /clone="Q11L10"
                     /clone_lib="LL21NCO2-Q Cosmid library"
                     /note="Accession No. AL109616"
     source          310004. .>340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.2"
                     /clone="pS11, 3' partial"
                     /clone_lib="CMP21 P1 library"
                     /note="Accession No. AP000223"
     repeat_region   2239. .2539
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(2607. .2851)
                     /note="MLT1A1"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   3128. .3427
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(5176. .5395)
                     /note="MIR"
                     /rpt_family="SINE/MIR"
                     /rpt_type=DISPERSED
     repeat_region   5492. .5603
                     /note="(TAGA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   5610. .5632
                     /note="(TGGA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   5633. .5654
                     /note="(ATG)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   5655. .5716
                     /note="(TGGA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   complement(5841. .5923)
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(6425. .6667)
                     /note="MIR"
                     /rpt_family="SINE/MIR"
                     /rpt_type=DISPERSED
     repeat_region   7339. .7494
                     /note="L1ME3A"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   7495. .7527
                     /note="(CA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   7528. .7775
                     /note="L1ME3A"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(7867. .8094)
                     /note="MER20"
                     /rpt_family="DNA/MER1_type"
                     /rpt_type=DISPERSED

  Query Match             10.7%;  Score 214;  DB 8;  Length 340000;
  Best Local Similarity   72.0%;  Pred. No. 2.5e-38;
  Matches  321;  Conservative    0;  Mismatches  120;  Indels    5;  Gaps    3;

Qy       1167 TTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTT 1226
             ||  | ||  | |  ||||  |||||||     |||  | |  ||||  |    | |||||  |||
Db     241613 TTGAATTGATAATTTTGTCCAGAGAAAACCTTGGATTTTTTTTTTTTTTTTTTTTTTTTTT 241554

Qy       1227 TTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGT 1286
             ||  ||||||||||  |  |||||||  ||  ||||||  ||||||||||  ||  |||  ||  |  |
Db     241553 TTTTTTTTTTTGAGAAGGAGTCTCGCTCTGTCACTCAGGCTGGAGTGCATTGG-TGCGAT 241495

Qy       1287 CTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCG 1344
             ||||||||||||||||||||||||||  ||||||||||||||  |||  |||||  |||  |
Db     241494 CTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTTCTGCCTCAGCCTCCTG 241435

Qy       1345 AATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGA 1402
             |  |||||||||  ||||||  ||  |||   |   |    |   ||||  ||||||||||||||||||
Db     241434 AGTAGCTGGGACTACAGGAGCGTGCCACCACTCCCAGCTAACTTTTGTATTTTTAGTAGA 241375

Qy       1403 GACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACC 1462
             ||  ||  |||||  ||||  |||  |||  ||  |  ||  ||||||||||  |||||||||  ||||
Db     241374 GAAAGGGTTTCACCATATTGGTCAGTCTGGTCTCGAACTCCTGACCTCAGGTGATGCACC 241315

Qy       1463 CACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTT 1522
             ||||||||||||||||||  ||||||  ||||||||||||||||||  |||||  ||  |  |  |
Db     241314 CACCTCAGCCTCCCAAAGTGCTGGGACTACAGGCATGAGCCACCATGCCCGGCTGGATAT 241255

Qy       1523 CTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTCATTTTGGAAGT 1582
               |  |  |  |  |||||  ||  |  ||  ||  |  |  |  |  ||  ||||
```

```
Db      241254 TTAACTTATTATATTTTATTGCTAGCTATTAATGCTTTCTATATGCTAAGCATTCAAAGT 241195

Qy        1583 AGTACACCTAAGTAAGGTTTTTAACA 1608
               |  ||  |  | |  |  |||| |
Db      241194 CATTTACAAATATTAACTGCTTAAAA 241169


RESULT 21
AC110054/c
LOCUS       AC110054               164857 bp    DNA     linear   PRI 01-MAY-2002
DEFINITION  Homo sapiens chromosome 15, clone CTD-2116N17, complete sequence.
ACCESSION   AC110054
VERSION     AC110054.4  GI:20377046
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 164857)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone CTD-2116N17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164857)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 164857)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 164857)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 1, 2002 this sequence version replaced gi:19033639.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23190
                Center clone name: 2116_N_17
            --------------

            Only the first 164.9 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC100840 [WICGR project
            L21571].
FEATURES             Location/Qualifiers
     source          1. .164857
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="CTD-2116N17"
                     /clone_lib="CITD1 Human BAC"
     repeat_region   complement(233. .311)
                     /rpt_family="MER5A"
     repeat_region   644. .664
                     /rpt_family="AT_rich"
     repeat_region   complement(665. .962)
                     /rpt_family="AluY"
     repeat_region   963. .980
                     /rpt_family="AT_rich"
     repeat_region   complement(1216. .1519)
                     /rpt_family="AluSx"
     repeat_region   complement(2545. .2842)
                     /rpt_family="AluJo"
     repeat_region   2845. .3183
                     /rpt_family="L2"
     repeat_region   3308. .3331
                     /rpt_family="(A)n"
     repeat_region   complement(3384. .3686)
                     /rpt_family="AluSq"
     repeat_region   complement(3691. .3848)
```

```
                     /rpt_family="L2"
     repeat_region   3849. .4144
                     /rpt_family="AluJb"
     repeat_region   complement(4145. .4298)
                     /rpt_family="L2"
     repeat_region   complement(4326. .4545)
                     /rpt_family="L2"
     repeat_region   4918. .5042
                     /rpt_family="MER85"
     repeat_region   complement(5204. .5319)
                     /rpt_family="AluY"
     repeat_region   complement(5334. .5640)
                     /rpt_family="AluSq"
     repeat_region   5896. .6188
                     /rpt_family="AluSx"
     repeat_region   6298. .6332
                     /rpt_family="AT_rich"
     repeat_region   complement(6346. .6633)
                     /rpt_family="MER1B"
     repeat_region   complement(6664. .6926)
                     /rpt_family="L1ME"
     repeat_region   complement(7320. .7626)
                     /rpt_family="AluSc"
     repeat_region   complement(8043. .8352)
                     /rpt_family="AluY"
     repeat_region   complement(8660. .8962)
                     /rpt_family="AluSq"
     repeat_region   9256. .9442
                     /rpt_family="AluSq"
     repeat_region   9443. .9504
                     /rpt_family="AluSq"
     repeat_region   9505. .9539
                     /rpt_family="(TTAAA)n"
     repeat_region   complement(9567. .10710)
                     /rpt_family="L1PB3"
     unsure          10683. .10783
                     /note="single clone coverage"
     repeat_region   complement(11072. .11376)
                     /rpt_family="AluSx"
     repeat_region   11660. .11671
                     /rpt_family="(CA)n"
     repeat_region   11672. .11706
                     /rpt_family="(CA)n"
     repeat_region   11737. .11764
                     /rpt_family="AT_rich"
     repeat_region   complement(11922. .12167)
                     /rpt_family="HAL1"
     repeat_region   complement(12249. .12438)
                     /rpt_family="HAL1"
     repeat_region   12445. .12977
                     /rpt_family="L3"
     repeat_region   complement(12984. .13117)
                     /rpt_family="MIR3"
     repeat_region   complement(13565. .13678)
                     /rpt_family="L2"
     repeat_region   15124. .15145
                     /rpt_family="AT_rich"
     repeat_region   15499. .15524
                     /rpt_family="AT_rich"
     repeat_region   15666. .16484
                     /rpt_family="L1MB6"
     repeat_region   16487. .16620
                     /rpt_family="AluSq/x"

  Query Match             10.7%;  Score 213.8;  DB 8;  Length 164857;
  Best Local Similarity   79.1%;  Pred. No. 3.1e-38;
  Matches  292;  Conservative    0;  Mismatches   72;  Indels    5;  Gaps    3;

Qy       1182 TGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAA 1241
             ||| || ||| || ||||     ||||     |    |||| ||||| |||||||||| |
Db      30813 TGTGCACAGAGGTGTTTGTAGCAGTTTCTGATGGCTATTTTTTTTTTTTTTTTTTTGAGA 30754

Qy       1242 CAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAA 1301
```

```
             | ||||||  || |||  |||||||||||| |||||| || | |||| ||||||||||
Db      30753 CGGAGTCTTGCTCTGTGGCCCAGGCTGGAGTGCAGTGG-TGCGATCTCAGCTCACTGCAA 30695

Qy       1302 CCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTAC 1359
             |||||||||||| |||||||||||||||||||  ||||| |||||| ||| |||||||||
Db      30694 CCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGGTGGGATTAC 30635

Qy       1360 AGGCGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCA 1417
             ||| |||| |  |  | | |  |||||||||||||||||||| |||||||| |||| ||||
Db      30634 AGGTGCATACTACCACGTCCAGCTAATTTTTGTATTTTTAGCAGAGACAGGGTTTTGCCA 30575

Qy       1418 TGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCA 1477
             ||||| |||| || | |||||||||||||| |||||||||| ||||||||||||||||||
Db      30574 TGTTGGCCAGACTGGTCTTGAACTCCTGACCTCAGGTGATCCGCCCACCTCAGCCTCCCA 30515

Qy       1478 AAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATT 1537
             |||  |||||||||||||  ||||||||| ||||| |||| |  |  |   |  | | ||
Db      30514 AAGTGCTGGGATTACAGGTGTGAGCCACCATGCCCGGCCTCTGATGGCTATTTTTATTTT 30455

Qy       1538 TGTTTCTGA 1546
             | ||| |||
Db      30454 TATTTTTGA 30446


RESULT 22
AR659559/c
LOCUS       AR659559               72048 bp    DNA     linear   PAT 13-JUN-2005
DEFINITION  Sequence 600 from patent US 6900016.
ACCESSION   AR659559
VERSION     AR659559.1  GI:67595583
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 72048)
  AUTHORS   Venter,J.C., Zhang,J.N., Liu,X., Rowe,W., Cravchik,A., Kalush,F.,
            Naik,A., Subramanian,G. and Woodage,T.
  TITLE     Polymorphisms in known genes associated with inflammatory
            autoimmune disease, methods of detection and uses thereof
  JOURNAL   Patent: US 6900016-A 600 31-MAY-2005;
            Applera Corporation; Norwalk, CT
FEATURES             Location/Qualifiers
     source          1. .72048
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             10.7%;  Score 213.6;  DB 6;  Length 72048;
  Best Local Similarity   74.0%;  Pred. No. 3.9e-38;
  Matches  311;  Conservative    0;  Mismatches  104;  Indels    5;  Gaps    3;

Qy       1208 TTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCT 1267
             | ||   || | ||| | ||| |||||||||| ||  |||||| || ||||||||||||||
Db      64752 TACTTGTATTTCTTTTTCTTTTTTTTTTTTTGAGACGAAGTCTCTCTCTGTCACCCAGGCT 64693

Qy       1268 GGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 1327
             |||||  |||||||    | |||| ||||||||||||||||||||| ||||||||||||||
Db      64692 GGAGTGCAGTGGC-ACGATCTCAGCTCACTGCAACCTCTGCCTCTCAGGTTCAAGCGATT 64634

Qy       1328 CTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAA 1383
             ||||||  | ||| |||||| |||||||||||||||| ||  ||  |  ||  || |||
Db      64633 CTCCTGCCTTAGCCTCCCGAGTAGCTGGGATTACAGGTGCCAGCCACCACACCTGGCTAC 64574

Qy       1384 TTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCC 1443
             |||||||||||||||||| ||||| |||||||||||||  |||||| | ||||||||||
Db      64573 TTTTTGTATTTTTAGTAGGGACAGGGTTTCGCCATGTTGGTCAGGCTGGTCTTGAACTCC 64514

Qy       1444 TGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCC 1503
             ||||  ||||||||||||||||||||  ||||||||||  |||||||  |||||||||| ||
Db      64513 TGACCTCAGGTGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATAACAGGCATGAACC 64454

Qy       1504 ACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
             |||||||||  |||   | |   |  || | | ||   ||           ||||   |||
```

```
Db      64453 ACCGTGCCCCGCCATGTCTTGTATTTCTTTTGTTAAATTTATTCCTAAGTATTTTATTCT 64394

Qy       1564 TCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGG 1623
              | ||| || | ||| || | | |    ||| |   | |||    ||    ||| | | |
Db      64393 TTTTAATCCTATTGTAAATGGGATTTTTAAAAATTATGTTTTTGGATTGTTTATTTCTAG 64334


RESULT 23
AR659643/c
LOCUS       AR659643               72048 bp    DNA     linear   PAT 13-JUN-2005
DEFINITION  Sequence 684 from patent US 6900016.
ACCESSION   AR659643
VERSION     AR659643.1  GI:67595683
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 72048)
  AUTHORS   Venter,J.C., Zhang,J.N., Liu,X., Rowe,W., Cravchik,A., Kalush,F.,
            Naik,A., Subramanian,G. and Woodage,T.
  TITLE     Polymorphisms in known genes associated with inflammatory
            autoimmune disease, methods of detection and uses thereof
  JOURNAL   Patent: US 6900016-A 684 31-MAY-2005;
            Applera Corporation; Norwalk, CT
FEATURES             Location/Qualifiers
     source          1. .72048
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             10.7%;  Score 213.6;  DB 6;  Length 72048;
  Best Local Similarity   74.0%;  Pred. No. 3.9e-38;
  Matches  311;  Conservative    0;  Mismatches  104;  Indels    5;  Gaps    3;

Qy       1208 TTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCT 1267
              | ||   || | ||| | ||| |||||||||| ||  |||||| || ||||||||||||
Db      64752 TACTTGTATTTCTTTTTCTTTTTTTTTTTTGAGACGAAGTCTCTCTCTGTCACCCAGGCT 64693

Qy       1268 GGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 1327
              ||||| |||||||    | |||| ||||||||||||||||||||| ||||||||||||||
Db      64692 GGAGTGCAGTGGC-ACGATCTCAGCTCACTGCAACCTCTGCCTCTCAGGTTCAAGCGATT 64634

Qy       1328 CTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAA 1383
              ||||||  | ||| |||||| |||||||||||||||| ||  ||  |  ||  || |||
Db      64633 CTCCTGCCTTAGCCTCCCGAGTAGCTGGGATTACAGGTGCCAGCCACCACACCTGGCTAC 64574

Qy       1384 TTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCC 1443
              |||||||||||||||| ||||| |||||||||||||  |||||| | |||||||||
Db      64573 TTTTTGTATTTTTAGTAGGGACAGGGTTTCGCCATGTTGGTCAGGCTGGTCTTGAACTCC 64514

Qy       1444 TGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCC 1503
              |||| |||||||||||||||||  |||||||||||  |||||||  |||||||||| ||
Db      64513 TGACCTCAGGTGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATAACAGGCATGAACC 64454

Qy       1504 ACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
              |||||||||| |||    | |    |  || | | ||   ||           ||||  |||
Db      64453 ACCGTGCCCCGCCATGTCTTGTATTTCTTTTGTTAAATTTATTCCTAAGTATTTTATTCT 64394

Qy       1564 TCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGG 1623
              | ||| || | ||| || | | |    ||| |   | |||    ||    ||| | | |
Db      64393 TTTTAATCCTATTGTAAATGGGATTTTTAAAAATTATGTTTTTGGATTGTTTATTTCTAG 64334


RESULT 24
AC068969/c
LOCUS       AC068969              168751 bp    DNA     linear   HTG 07-MAY-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-185A14 map 15, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC068969
VERSION     AC068969.3  GI:13959305
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 168751)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-185A14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168751)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 7, 2001 this sequence version replaced gi:8247910.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9836
                Center clone name: 185_A_14
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 91% of reads
                Sequencing vector: Plasmid; n/a; 9% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 162617 bases at least Q40
                Consensus quality: 165848 bases at least Q30
                Consensus quality: 167104 bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: 167651; sum-of-contigs
                Quality coverage: 5.4 in Q20 bases; agarose-fp
                Quality coverage: 5.5 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5645: contig of 5645 bp in length
            *     5646     5745: gap of 100 bp
            *     5746     6917: contig of 1172 bp in length
            *     6918     7017: gap of 100 bp
            *     7018     8320: contig of 1303 bp in length
            *     8321     8420: gap of 100 bp
            *     8421    10553: contig of 2133 bp in length
            *    10554    10653: gap of 100 bp
            *    10654    13543: contig of 2890 bp in length
```

```
          *    13544    13643: gap of 100 bp
          *    13644    19241: contig of 5598 bp in length
          *    19242    19341: gap of 100 bp
          *    19342    35264: contig of 15923 bp in length
          *    35265    35364: gap of 100 bp
          *    35365    68938: contig of 33574 bp in length
          *    68939    69038: gap of 100 bp
          *    69039    90272: contig of 21234 bp in length
          *    90273    90372: gap of 100 bp
          *    90373   113565: contig of 23193 bp in length
          *   113566   113665: gap of 100 bp
          *   113666   146203: contig of 32538 bp in length
          *   146204   146303: gap of 100 bp
          *   146304   168751: contig of 22448 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168751
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-185A14"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .5645
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             5646. .5745
                     /estimated_length=100
     misc_feature    5746. .6917
                     /note="assembly_fragment"
     gap             6918. .7017
                     /estimated_length=100
     misc_feature    7018. .8320
                     /note="assembly_fragment"
     gap             8321. .8420
                     /estimated_length=100
     misc_feature    8421. .10553
                     /note="assembly_fragment"
     gap             10554. .10653
                     /estimated_length=100
     misc_feature    10654. .13543
                     /note="assembly_fragment"
     gap             13544. .13643
                     /estimated_length=100
     misc_feature    13644. .19241
                     /note="assembly_fragment"
     gap             19242. .19341
                     /estimated_length=100
     misc_feature    19342. .35264
                     /note="assembly_fragment"
     gap             35265. .35364
                     /estimated_length=100
     misc_feature    35365. .68938
                     /note="assembly_fragment"
     gap             68939. .69038
                     /estimated_length=100
     misc_feature    69039. .90272
                     /note="assembly_fragment"
     gap             90273. .90372
                     /estimated_length=100
     misc_feature    90373. .113565
                     /note="assembly_fragment"
     gap             113566. .113665
                     /estimated_length=100
     misc_feature    113666. .146203
                     /note="assembly_fragment"
     gap             146204. .146303
                     /estimated_length=100
     misc_feature    146304. .168751
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
```

```
ORIGIN

  Query Match           10.7%;  Score 213.6;  DB 14;  Length 168751;
  Best Local Similarity   74.0%;  Pred. No. 3.5e-38;
  Matches  311;  Conservative    0;  Mismatches  104;  Indels    5;  Gaps    3;

Qy       1208 TTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCT 1267
             | ||   || | ||| | ||| |||||||||| ||  |||||| || ||||||||||||
Db      47984 TACTTGTATTTCTTTTTCTTTTTTTTTTTTGAGACGAAGTCTCTCTCTGTCACCCAGGCT 47925

Qy       1268 GGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 1327
             ||||| |||||||   | |||| ||||||||||||||||||||| ||||||||||||||||
Db      47924 GGAGTGCAGTGGC-ACGATCTCAGCTCACTGCAACCTCTGCCTCTCAGGTTCAAGCGATT 47866

Qy       1328 CTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAA 1383
             ||||||  | ||| |||||| |||||||||||||||| ||  ||  |  ||  || |||
Db      47865 CTCCTGCCTTAGCCTCCCGAGTAGCTGGGATTACAGGTGCCAGCCACCACACCTGGCTAC 47806

Qy       1384 TTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCC 1443
             |||||||||||||||||| ||||| |||||||||||||  |||||| | ||||||||||
Db      47805 TTTTTGTATTTTTAGTAGGGACAGGGTTTCGCCATGTTGGTCAGGCTGGTCTTGAACTCC 47746

Qy       1444 TGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCC 1503
             |||| ||||||||||||||||||  ||||||||||  |||||| |||||||||| ||
Db      47745 TGACCTCAGGTGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATAACAGGCATGAACC 47686

Qy       1504 ACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
             |||||||||| |||   | |   |  || | | ||   ||           ||||   |||
Db      47685 ACCGTGCCCCGCCATGTCTTGTATTTCTTTTGTTAAATTTATTCCTAAGTATTTTATTCT 47626

Qy       1564 TCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGG 1623
             | ||| || | ||| || | | |    ||| |   | |||    ||    ||| | | |
Db      47625 TTTTAATCCTATTGTAAATGGGATTTTTAAAAATTATGTTTTTGGATTGTTTATTTCTAG 47566


RESULT 25
AC067805/c
LOCUS       AC067805              170883 bp    DNA     linear   PRI 25-NOV-2001
DEFINITION  Homo sapiens chromosome 15, clone RP11-63A23, complete sequence.
ACCESSION   AC067805
VERSION     AC067805.9  GI:17061076
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 170883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-63A23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 170883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 23, 2001 this sequence version replaced gi:16554401.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9837
                Center clone name: 63_A_23
            --------------.
FEATURES             Location/Qualifiers
     source          1. .170883
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-63A23"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(427. .489)
                     /rpt_family="L2"
     repeat_region   520. .594
                     /rpt_family="MIR"
     repeat_region   895. .1068
                     /rpt_family="MIR"
     repeat_region   complement(1233. .1527)
                     /rpt_family="AluSx"
     repeat_region   complement(1707. .1826)
                     /rpt_family="MER112"
     repeat_region   1847. .1918
                     /rpt_family="(TA)n"
     repeat_region   2628. .2666
                     /rpt_family="(TG)n"
     repeat_region   complement(3097. .3216)
                     /rpt_family="MIR"
     repeat_region   complement(3246. .3375)
                     /rpt_family="L2"
     repeat_region   3414. .3652
                     /rpt_family="MER20"
     repeat_region   3811. .4004
                     /rpt_family="MIR"
     repeat_region   4208. .4301
                     /rpt_family="MIR"
     repeat_region   complement(4895. .5001)
                     /rpt_family="MIR3"
     repeat_region   complement(5600. .5751)
                     /rpt_family="MER5B"
     repeat_region   7380. .7697
                     /rpt_family="L2"
     repeat_region   7965. .8075
                     /rpt_family="L2"
     repeat_region   9003. .9162
                     /rpt_family="MIR"
     repeat_region   9475. .9563
                     /rpt_family="MER91C"
     repeat_region   10157. .10454
                     /rpt_family="AluSx"
     repeat_region   10537. .10563
                     /rpt_family="(A)n"
     repeat_region   complement(10705. .11175)
                     /rpt_family="MLT1C"
     repeat_region   complement(11404. .11516)
                     /rpt_family="MIR"
     repeat_region   complement(12012. .12321)
                     /rpt_family="AluY"
     repeat_region   12452. .12757
                     /rpt_family="AluJb"
     repeat_region   complement(12767. .12872)
                     /rpt_family="L1ME4A"
     repeat_region   complement(12988. .13085)
                     /rpt_family="L2"
     repeat_region   13670. .13826
                     /rpt_family="L1MB3"
     repeat_region   14301. .14334
                     /rpt_family="L2"
     repeat_region   14361. .14652
                     /rpt_family="AluJo"
     repeat_region   14654. .14678
                     /rpt_family="AT_rich"
     repeat_region   14787. .14997
                     /rpt_family="MIR"
     repeat_region   15459. .15563
                     /rpt_family="L2"
     repeat_region   complement(15651. .15701)
                     /rpt_family="AluJ"
     repeat_region   complement(15702. .16075)
                     /rpt_family="THE1C"
     repeat_region   complement(16076. .16202)
                     /rpt_family="AluJ"
     repeat_region   complement(16249. .16331)
                     /rpt_family="AluJ"
     repeat_region   complement(16668. .17193)
                     /rpt_family="MER77"
     repeat_region   complement(17350. .17613)
                     /rpt_family="AluSx"
     repeat_region   18923. .19228
                     /rpt_family="AluY"
     repeat_region   complement(19290. .19476)
                     /rpt_family="L2"
     repeat_region   19477. .19781
                     /rpt_family="AluSx"
     repeat_region   complement(19782. .20084)
                     /rpt_family="L2"
     repeat_region   21246. .21273
                     /rpt_family="(T)n"
     repeat_region   21770. .21894
                     /rpt_family="MIR"
     repeat_region   complement(22641. .22815)
                     /rpt_family="MIR"
     repeat_region   complement(24142. .24281)
                     /rpt_family="MIR"
     repeat_region   complement(25199. .25875)
                     /rpt_family="MER7C"
     repeat_region   25977. .26089
```

```
                     /rpt_family="MER63A"
     repeat_region   complement(26400. .26503)
                     /rpt_family="L2"
     repeat_region   26523. .26693
                     /rpt_family="MIR3"
     repeat_region   26691. .27038
                     /rpt_family="L3"
     repeat_region   27208. .27471
                     /rpt_family="AluSx"
     repeat_region   complement(27615. .27663)
                     /rpt_family="L3"
     repeat_region   29661. .29770
                     /rpt_family="MER5B"
     repeat_region   29844. .29957
                     /rpt_family="MER91B"
     repeat_region   complement(30204. .30242)
                     /rpt_family="MLT2B1"
     repeat_region   30243. .30302

  Query Match             10.7%;  Score 213.6;  DB 8;  Length 170883;
  Best Local Similarity   74.0%;  Pred. No. 3.4e-38;
  Matches  311;  Conservative    0;  Mismatches  104;  Indels    5;  Gaps    3;

Qy       1208 TTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCT 1267
             | ||    || | ||| | ||| |||||||||| ||   |||||| || ||||||||||||
Db     111933 TACTTGTATTTCTTTTTCTTTTTTTTTTTTTGAGACGAAGTCTCTCTCTGTCACCCAGGCT 111874

Qy       1268 GGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 1327
             ||||| |||||||   | |||| ||||||||||||||||||||| ||||||||||||||
Db     111873 GGAGTGCAGTGGC-ACGATCTCAGCTCACTGCAACCTCTGCCTCTCAGGTTCAAGCGATT 111815

Qy       1328 CTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAA 1383
             ||||||  | ||| |||||| |||||||||||||||| ||  ||  |  ||  || |||
Db     111814 CTCCTGCCTTAGCCTCCCGAGTAGCTGGGATTACAGGTGCCAGCCACCACACCTGGCTAC 111755

Qy       1384 TTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCC 1443
             |||||||||||||||| ||||| |||||||||||||  |||||| | ||||||||||
Db     111754 TTTTTGTATTTTTAGTAGGGACAGGGTTTCGCCATGTTGGTCAGGCTGGTCTTGAACTCC 111695

Qy       1444 TGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCC 1503
             |||| |||||||||||||||||  |||||||||| |||||| |||||||||| ||
Db     111694 TGACCTCAGGTGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATAACAGGCATGAACC 111635

Qy       1504 ACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
             ||||||||| |||   | |   |  || | | ||   ||           ||||   |||
Db     111634 ACCGTGCCCCGCCATGTCTTGTATTTCTTTTGTTAAATTTATTCCTAAGTATTTTATTCT 111575

Qy       1564 TCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGG 1623
             | ||| || | ||| || | | |    ||| |   | |||    ||    ||| | | |
Db     111574 TTTTAATCCTATTGTAAATGGGATTTTTAAAAATTATGTTTTTGGATTGTTTATTTCTAG 111515


RESULT 26
AC091721
LOCUS       AC091721               175752 bp    DNA     linear   HTG 23-MAY-2002
DEFINITION  Pan troglodytes clone RP43-97L19, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC091721
VERSION     AC091721.2  GI:21104890
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 175752)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175752)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 175752)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On May 23, 2002 this sequence version replaced gi:14192901.
            ------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ------------- Project Information
                 Center project name: cea
                 Center clone name: 097L19

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 174154 bases at least Q40
                 Consensus quality: 174933 bases at least Q30
                 Consensus quality: 175303 bases at least Q20
                 Insert size: 166000; agarose-fp
                 Insert size: 175452; sum-of-contigs
                 Quality coverage: 14.26x in Q20 bases; agarose-fp
                 Quality coverage: 13.49x in Q20 bases; sum-of-contigs
            -------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    68459: contig of 68459 bp in length
            *   68460    68559: gap of unknown length
            *   68560   152280: contig of 83721 bp in length
            *  152281   152380: gap of unknown length
            *  152381   166826: contig of 14446 bp in length
            *  166827   166926: gap of unknown length
            *  166927   175752: contig of 8826 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175752
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-97L19"
                     /clone_lib="RP43"
     misc_feature    1. .68459
```

```
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    1. .36692
                     /note="clone overlaps with GenBank Accession Number
                     AC091717 clone RP43-184N1 (center project name aoz)"
     gap             68460. .68559
                     /estimated_length=unknown
     misc_feature    68560. .152280
                     /note="assembly_fragment"
     misc_feature    131176. .175752
                     /note="clone overlaps with GenBank Accession Number
                     AC091716 clone RP43-147G10 (center project name ceb)"
     gap             152281. .152380
                     /estimated_length=unknown
     misc_feature    152381. .166826
                     /note="assembly_fragment"
     gap             166827. .166926
                     /estimated_length=unknown
     misc_feature    166927. .175752
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             10.7%;  Score 213.4;  DB 14;  Length 175752;
  Best Local Similarity   61.4%;  Pred. No. 3.8e-38;
  Matches  412;  Conservative    0;  Mismatches  251;  Indels    8;  Gaps    4;

Qy       1205 GTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAG 1264
Db     127196 GCTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATAGACCGAGTCTTGCTCTGTCACTCAG 127255

Qy       1265 GCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCG 1324
Db     127256 GCTGGAGTGCAGTGGC-GCGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCG 127314

Qy       1325 ATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG----CACCACCATG 1378
Db     127315 ATTCTCCTGCCTCAGCCTCCCGAATAGCTGGGATTACAGGCGCTTGCCATCATACACCCA 127374

Qy       1379 CCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGA 1438
Db     127375 GCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGA 127434

Qy       1439 ACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCAT 1498
Db     127435 ACTCCTGACCTCAGGTGATCTGCCTGCCTTGGCCTCCCAGAGTGCTGGGATTACAGGCGT 127494

Qy       1499 GAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTC 1558
Db     127495 GAGCCACTGCGCCCGGCCCCAGCTCTTGTTTCTTTTTTTTTTTTAATTAGAAATAACTGT 127554

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618
Db     127555 GTTGTTTCTTTTGA-TTAGAAATAACTGTGGTGTTTCTAATTGTATTTCTTATAACTGTA 127613

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678
Db     127614 TTTCTGTGTTTAGTGATTACTTTTATAATTTTATAGAATACCTCTTATTTTTAGATACTG 127673

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738
Db     127674 TTTTTTCTATGAGCAATAATGAAATTAGTGTATTTCCAATTTTTCTCCTTCCTTCCCGTG 127733

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798
Db     127734 TTCTAATTTTAGTTGTTACATTACCTAATGTTTGCTTTTATACTATTAAGTTTTTTTTTG 127793

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858
Db     127794 TACTTCTCTTTGCTTTGTTAGCTTTAAATAATACTTTTTGACACCTAGTTATTTATATCA 127853

Qy       1859 ACTGCATGCCA 1869
Db     127854 TCTTCCCCCCA 127864


RESULT 27
AC146040/c
LOCUS       AC146040               175808 bp    DNA     linear   PRI 03-JAN-2004
DEFINITION  Pan troglodytes BAC clone RP43-97N16 from 7, complete sequence.
ACCESSION   AC146040
VERSION     AC146040.2  GI:40254142
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 175808)
  AUTHORS   Harkins,R., Kozlowicz,A. and Bielicki,L.
  TITLE     The sequence of Pan troglodytes BAC clone RP43-97N16
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 175808)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Pan troglodytes
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 175808)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 175808)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 175808)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2004) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Dec 20, 2003 this sequence version replaced gi:33386992.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            ------------- Summary Statistics
                Center project name: C_PT097N16
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
            was isolated from white blood cells obtained from a male chimpanzee
            (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
```

clone and detailed information can be obtained from ResGen (http://www.resgen.com) or Pieter de Jong and co-workers at http://www.bacpac.chori.org.

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

```
FEATURES             Location/Qualifiers
     source          1. .175808
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /map="7"
                     /clone="RP43-97N16"
                     /clone_lib="RPCI-43"
ORIGIN
```

Query Match 10.7%; Score 213.4; DB 8; Length 175808;
Best Local Similarity 61.4%; Pred. No. 3.8e-38;
Matches 412; Conservative 0; Mismatches 251; Indels 8; Gaps 4;

```
Qy       1205 GTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAG 1264
Db      48367 GCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATAGACCGAGTCTTGCTCTGTCACTCAG 48308

Qy       1265 GCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCG 1324
Db      48307 GCTGGAGTGCAGTGGC-GCGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCG 48249

Qy       1325 ATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG----CACCACCATG 1378
Db      48248 ATTCTCCTGCCTCAGCCTCCCGAATAGCTGGGATTACAGGCGCTTGCCATCATACACCCA 48189

Qy       1379 CCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGA 1438
Db      48188 GCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGA 48129

Qy       1439 ACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCAT 1498
Db      48128 ACTCCTGACCTCAGGTGATCTGCCTGCCTTGGCCTCCCAGAGTGCTGGGATTACAGGCGT 48069

Qy       1499 GAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTC 1558
Db      48068 GAGCCACTGCGCCCGGCCCCAGCTCTTGTTTCTTTTTTTTTTTTAATTAGAAATAACTGT 48009

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618
Db      48008 GTTGTTTCTTTTGA-TTAGAAATAACTGTGGTGTTTCTAATTGTATTTCTTATAACTGTA 47950

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678
Db      47949 TTTCTGTGTTTAGTGATTACTTTTATAATTTTATAGAATACCTCTTATTTTTAGATACTG 47890

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738
Db      47889 TTTTTTCTATGAGCAATAATGAAATTAGTGTATTTCCAATTTTTCTCCTTCCTTCCCGTG 47830

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798
Db      47829 TTCTAATTTTAGTTGTTACATTACCTAATGTTTGCTTTTATACTATTAAGTTTTTTTTTG 47770

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858
Db      47769 TACTTCTCTTTGCTTTGTTAGCTTTAAATAATACTTTTTGACACCTAGTTATTTATATCA 47710

Qy       1859 ACTGCATGCCA 1869
Db      47709 TCTTCCCCCCA 47699
```

RESULT 28
AC138517/c
LOCUS AC138517 42673 bp DNA linear PRI 25-MAR-2003

```
DEFINITION  Homo sapiens chromosome 5 clone XXfos-850G5, complete sequence.
ACCESSION   AC138517
VERSION     AC138517.2  GI:29171394
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 42673)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 42673)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 42673)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 25, 2003 this sequence version replaced gi:27545018.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .42673
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="XXfos-850G5"
ORIGIN
```

Query Match 10.7%; Score 213.2; DB 8; Length 42673;
Best Local Similarity 78.2%; Pred. No. 5.1e-38;
Matches 294; Conservative 0; Mismatches 78; Indels 4; Gaps 3;

```
Qy       1204 TGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCA 1263
Db       1391 TTTTTTCTGTTTTCTTTTCTTTTTTTTTTTTTTGAGACAGAGTCTCACTTTGTCTCCCA 1332

Qy       1264 GGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGC 1323
Db       1331 GGCTGGAGTGCAATGGC-ATGATCTCAGCTCACTGCAACCTCTACCTCCCGGGTTCAAGC 1273

Qy       1324 GATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG-CACCACCATGCC 1380
Db       1272 GATTCTCATGCCTCAGCCTCCCAAGCAGCTGGGATTACAGGCGTGTGCCATCAGGCCTGC 1213

Qy       1381 TAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAAC 1440
Db       1212 TAATTTTTGTATTTTTAATAGAGGCAGGGTTTCACCATGTTGGCCGGGCTAGTCTTGAAC 1153

Qy       1441 TCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGA 1500
Db       1152 TCCTGACCTCAAGTGATCTACCCACCTCAGCCTCCCTAAGTGCTGGGATTACAGGTGTGA 1093

Qy       1501 GCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTA 1560
Db       1092 ACCACCGCACCTGGCCAGTGTTTTCTTTTCTTTTTTCTTTTTCTTTCTTTTTTTTTTTT 1033

Qy       1561 TCTTCTTATTCATTTT 1576
Db       1032 TTTTTTTTTTTTTTT 1017
```

RESULT 29

```
AL590822
LOCUS       AL590822               179836 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-181G12 on chromosome 1 Contains
            the 3' end of the PRKCZ gene for protein kinase C zeta, a novel
            gene, a novel gene (FLJ31031), the SKI gene for v-ski sarcoma viral
            oncogene homolog (avian) and five CpG islands, complete sequence.
ACCESSION   AL590822
VERSION     AL590822.36  GI:32251606
KEYWORDS    HTG; FLJ1031; PRKCZ; SKI.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 179836)
  AUTHORS   Hall,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 25, 2003 this sequence version replaced gi:31071456.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-181G12 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            -------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .179836
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-181G12"
                     /clone_lib="RPCI-11.1"
     misc_feature    1
                     /note="Clone_left_end: RP11-181G12"
     gene            join(AL391845.49:90455. .90686,AL391845.49:95426. .95547,
                     AL391845.49:96469. .96558,AL391845.49:99526. .99576,
                     AL645703.18:44363. .44448,3076. .3207,4893. .4974,7738. .7790,
                     9656. .9844,14861. .14958,28384. .28470,30921. .31056,
                     31167. .31254,32763. .32882,33620. .33699,34090. .34179,
                     43449. .43564,43788. .44261)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-001"
     mRNA            join(AL391845.49:90455. .90686,AL391845.49:95426. .95547,
                     AL391845.49:96469. .96558,AL391845.49:99526. .99576,
                     AL645703.18:44363. .44448,3076. .3207,4893. .4974,7738. .7790,
                     9656. .9844,14861. .14958,28384. .28470,30921. .31056,
                     31167. .31254,32763. .32882,33620. .33699,34090. .34179,
                     43449. .43564,43788. .44261)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-001"
                     /product="protein kinase C, zeta"
                     /note="match: cDNAs: Em:BC008058.1 Em:BC008058.2
                     Em:BC014270.1 Em:BC014270.2 Em:BT007082.1 Em:L14283.1"
     CDS             join(AL391845.49:90616. .90686,AL391845.49:95426. .95547,
                     AL391845.49:96469. .96558,AL391845.49:99526. .99576,
                     AL645703.18:44363. .44448,3076. .3207,4893. .4974,7738. .7790,
                     9656. .9844,14861. .14958,28384. .28470,30921. .31056,
                     31167. .31254,32763. .32882,33620. .33699,34090. .34179,
                     43449. .43564,43788. .43875)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-001"
                     /standard_name="OTTHUMP00000001368"
                     /note="match: proteins: Sw:O19111 Sw:P09217 Sw:Q02956
                     Sw:Q05513 Tr:AAP35745 Tr:Q9JLP6"
                     /codon_start=1
                     /product="protein kinase C, zeta"
                     /protein_id="CAI12484.1"
                     /db_xref="GI:55961168"
                     /translation="MPSRTGPKMEGSGGRVRLKAHYGGDIFITSVDAATTFEELCEEV
                     RDMCRLHQQHPLTLKWVDSEGDPCTVSSQMELEEAFRLARQCRDEGLIIHVFPSTPEQ
                     PGLPCPGEDKSIYRRGARRWRKLYRANGHLFQAKRFNRRAYCGQCSERIWGLARQGYR
                     CINCKLLVHKRCHGLVPLTCRKHMDSVMPSQEPPVDDKNEDADLPSEETDGIAYISSS
                     RKHDSIKDDSEDLKPVIDGMDGIKISQGLGLQDFDLIRVIGRGSYAKVLLVRLKKNDQ
                     IYAMKVVKKELVHDDEDIDWVQTEKHVFEQASSNPFLVGLHSCFQTTSRLFLVIEYVN
                     GGDLMFHMQRQRKLPEEHARFYAAEICIALNFLHERGIIYRDLKLDNVLLDADGHIKL
                     TDYGMCKEGLGPGDTTSTFCGTPNYIAPEILRGEEYGFSVDWWALGVLMFEMMAGRSP
                     FDIITDNPDMNTEDYLFQVILEKPIRIPRFLSVKASHVLKGFLNKDPKERLGCRPQTG
                     FSDIKSHAFFRSIDWDLLEKKQALPPFQPQITDDYGLDNFDTQFTSEPVQLTPDDEDA
                     IKRIDQSEFEGFEYINPLLLSTEESV"
     gene            join(AL391845.49:113629. .113914,AL645703.18:44363. .44448,
                     3076. .3207,4893. .4974,7738. .7790,9656. .9844,14861. .14958,
                     28384. .28470)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-009"
     mRNA            join(AL391845.49:113629. .113914,AL645703.18:44363. .44448,
                     3076. .3207,4893. .4974,7738. .7790,9656. .9844,14861. .14958,
                     28384. .28470)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-009"
                     /product="protein kinase C, zeta"
                     /note="match: ESTs: Em:AL538613.2 Em:BI550905.1"
     gene            join(AL645703.18:44091. .44187,AL645703.18:44363. .44448,
                     3076. .3207,4255. .4377,4893. .4974,7738. .7790,9656. .>9845)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-011"
     mRNA            join(AL645703.18:44091. .44187,AL645703.18:44363. .44448,
                     3076. .3207,4255. .4377,4893. .4974,7738. .7790,9656. .>9845)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-011"
                     /product="protein kinase C, zeta"
                     /note="match: ESTs: Em:CD251012.1"
     gene            join(AL645703.18:20755. .20918,AL645703.18:44363. .44448,
                     3076. .3207,4893. .4974,7738. .7790,9656. .>9844)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-010"
     mRNA            join(AL645703.18:20755. .20918,AL645703.18:44363. .44448,
                     3076. .3207,4893. .4974,7738. .7790,9656. .>9844)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-010"
                     /product="protein kinase C, zeta"
                     /note="match: ESTs: Em:AL044353.1 Em:BI560781.1"
     gene            join(AL645703.18:13817. .13996,AL645703.18:44363. .44448,
                     3076. .3207,4893. .4974,7738. .7790,9656. .>9736)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-003"
     mRNA            join(AL645703.18:13817. .13996,AL645703.18:44363. .44448,
                     3076. .3207,4893. .4974,7738. .7790,9656. .>9736)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-003"
                     /product="protein kinase C, zeta"
                     /note="match: ESTs: Em:AL542576.2 Em:AU137858.1
                     Em:AU138321.1 Em:AU138397.1 Em:AU138706.1 Em:AU138889.1
                     Em:BG571791.1"
```

```
     gene            join(AL645703.18:42826. .43101,AL645703.18:44363. .44448,
                     3076. .3207,4893. .4974,7738. .>7790)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-004"
     mRNA            join(AL645703.18:42826. .43101,AL645703.18:44363. .44448,
                     3076. .3207,4893. .4974,7738. .>7790)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-004"
                     /product="protein kinase C, zeta"
                     /note="match: ESTs: Em:BE537476.1 Em:BF125400.1"
     gene            join(AL391845.49:90470. .90686,AL391845.49:95426. .95547,
                     AL391845.49:99526. .99576,AL645703.18:44363. .44448,
                     3076. .3207,4893. .>4974)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-007"
     mRNA            join(AL391845.49:90470. .90686,AL391845.49:95426. .95547,
                     AL391845.49:99526. .99576,AL645703.18:44363. .44448,
                     3076. .3207,4893. .>4974)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-007"
                     /product="protein kinase C, zeta"
                     /note="match: ESTs: Em:BG329855.1"
     gene            join(AL645703.18:43914. .44187,AL645703.18:44363. .44448,
                     3076. .3207,4893. .>4964)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-006"
     mRNA            join(AL645703.18:43914. .44187,AL645703.18:44363. .44448,
                     3076. .3207,4893. .>4964)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-006"
                     /product="protein kinase C, zeta"
                     /note="match: ESTs: Em:BG898174.1"
     gene            join(3317. .3450,4893. .4974,7738. .7790,9656. .9844,
                     14861. .14958,28384. .28470,30921. .31056,31167. .31254,
                     32763. .32882,33620. .33699,34090. .34179,43449. .43564,
                     43788. .44261)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-002"
     mRNA            join(3317. .3450,4893. .4974,7738. .7790,9656. .9844,
                     14861. .14958,28384. .28470,30921. .31056,31167. .31254,
                     32763. .32882,33620. .33699,34090. .34179,43449. .43564,
                     43788. .44261)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-002"
                     /product="protein kinase C, zeta"
                     /note="match: cDNAs: Em:AK097627.1"
     misc_feature    14579. .14721
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-009"
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
     gene            join(34127. .34179,40698. .40780,43449. .43564,43788. .43982)
                     /gene="PRKCZ"
                     /locus_tag="RP11-181G12.1-012"
     mRNA            join(34127. .34179,40698. .40780,43449. .43564,43788. .43982)

  Query Match             10.6%;  Score 212.8;  DB 8;  Length 179836;
  Best Local Similarity   79.2%;  Pred. No. 5.2e-38;
  Matches  290;  Conservative    0;  Mismatches   72;  Indels    4;  Gaps    3;

Qy       1156 AAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGA 1215
              || ||  |   ||||   |  | ||  |||     |   ||||  || | |||| ||
Db      46107 AAAAGCAGGTGTTTCCAAGAAGCAAAGGTTTCTGAAGCCATTTTCTTTTTTTTTTTCTTT 46166

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
              |  |||| ||||| |||||||||| ||||||| | ||| |||  |||||||||||||| ||
Db      46167 TTCTTTTTTTTTTCTTTTTTTTGAGACAGAGTATTACTCTGTAGCCCAGGCTGGAGTGCA 46226

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
              ||||| | | ||||||||||||||||||||  |||||| |||| ||||||||||||||
Db      46227 GTGGC-GAGATCTCGGCTCACTGCAACCTCCACCTCCCGGGTTTAAGCGATTCTCCTGCC 46285

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC-CATGCCTAATTTTTGTAT 1392
              ||||| |||| | |||||||||||||||| | ||| |||| |  | ||||||||||||
Db      46286 TCAGCCTCCCCAGTAGCTGGGATTACAGGCACCTGCCCCACGCCCGGCTAATTTTTGTAT 46345

Qy       1393 TTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAG 1452
              |||||||||||| | ||||| |||||||  |||||||  | |||||||||||||| ||||
Db      46346 TTTTAGTAGAGACGGGGTTTCACCATGTTTGCCAGGCTGGTCTTGAACTCCTGACCTCAG 46405

Qy       1453 GTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCC 1512
              ||||||||||  ||||||||||||||||  ||||||||||||| ||||| |||| ||||
Db      46406 GTGATCCACCTGCCTCAGCCTCCCAAAGTTCTGGGATTACAGGCGTGAGCTACCGCGCCC 46465

Qy       1513 AGCCTG 1518
              ||||||
Db      46466 GGCCTG 46471


RESULT 30
AC068198
LOCUS       AC068198              213432 bp    DNA     linear   HTG 22-FEB-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-290B2, WORKING DRAFT SEQUENCE,
            10 unordered pieces.
ACCESSION   AC068198
VERSION     AC068198.11  GI:9930748
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 213432)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Lam,B., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
            Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,
            Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 213432)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Aug 27, 2000 this sequence version replaced gi:8567772.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 887
                 Center clone name: RP11-290B2
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513
                 Chemistry: Dye-primer; 0% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 205371 bases at least Q40
                 Consensus quality: 208312 bases at least Q30
                 Consensus quality: 209125 bases at least Q20
                 Insert size: 197597; agarose-fp
                 Insert size: 212532; sum-of-contigs
                 Quality coverage:  8.3x in Q20 bases; agarose-fp
                 Quality coverage:  7.7x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1608: contig of 1608 bp in length
            *     1609     1708: gap of unknown length
            *     1709     5437: contig of 3729 bp in length
            *     5438     5537: gap of unknown length
            *     5538     9381: contig of 3844 bp in length
            *     9382     9481: gap of unknown length
            *     9482    20244: contig of 10763 bp in length
            *    20245    20344: gap of unknown length
            *    20345    33266: contig of 12922 bp in length
            *    33267    33366: gap of unknown length
            *    33367    54346: contig of 20980 bp in length
            *    54347    54446: gap of unknown length
            *    54447    78355: contig of 23909 bp in length
            *    78356    78455: gap of unknown length
            *    78456   100213: contig of 21758 bp in length
            *   100214   100313: gap of unknown length
            *   100314   124910: contig of 24597 bp in length
            *   124911   125010: gap of unknown length
            *   125011   213432: contig of 88422 bp in length.
FEATURES             Location/Qualifiers
     source          1. .213432
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-290B2"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .1608
                     /note="assembly_name:Contig20"
     gap             1609. .1708
                     /estimated_length=unknown
     misc_feature    1709. .5437
                     /note="assembly_name:Contig21
                     clone_end:SP6
                     vector_side:left"
     gap             5438. .5537
                     /estimated_length=unknown
     misc_feature    5538. .9381
                     /note="assembly_name:Contig22"
     gap             9382. .9481
                     /estimated_length=unknown
     misc_feature    9482. .20244
                     /note="assembly_name:Contig23"
     gap             20245. .20344
                     /estimated_length=unknown
     misc_feature    20345. .33266
                     /note="assembly_name:Contig24"
     gap             33267. .33366
                     /estimated_length=unknown
     misc_feature    33367. .54346
                     /note="assembly_name:Contig25"
     gap             54347. .54446
                     /estimated_length=unknown
     misc_feature    54447. .78355
                     /note="assembly_name:Contig26"
     gap             78356. .78455
                     /estimated_length=unknown
     misc_feature    78456. .100213
                     /note="assembly_name:Contig27
                     clone_end:T7
                     vector_side:right"
     gap             100214. .100313
                     /estimated_length=unknown
     misc_feature    100314. .124910
                     /note="assembly_name:Contig28"
     gap             124911. .125010
                     /estimated_length=unknown
     misc_feature    125011. .213432
                     /note="assembly_name:Contig29"
ORIGIN
```

```
  Query Match           10.6%;  Score 212.8;  DB 14;  Length 213432;
  Best Local Similarity   79.2%;  Pred. No. 5.1e-38;
  Matches  290;  Conservative    0;  Mismatches   72;  Indels    4;  Gaps    3;

Qy       1156 AAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGA 1215
              || ||  |   ||||   |  | ||  |||     |   ||||  || | |||| ||
Db      70335 AAAAGCAGGTGTTTCCAAGAAGCAAAGGTTTCTGAAGCCATTTTCTTTTTTTTTTCTTT 70394

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
              |  |||| ||||| |||||||||| ||||||| | ||| |||  ||||||||||||| ||
Db      70395 TTCTTTTTTTTTTCTTTTTTTTGAGACAGAGTATTACTCTGTAGCCCAGGCTGGAGTGCA 70454

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
              ||||| | | ||||||||||||||||||||  |||||| |||| |||||||||||||
Db      70455 GTGGC-GAGATCTCGGCTCACTGCAACCTCCACCTCCCGGGTTTAAGCGATTCTCCTGCC 70513

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC-CATGCCTAATTTTTGTAT 1392
              ||||| |||| | |||||||||||||||| | ||| |||| |  | ||||||||||||
Db      70514 TCAGCCTCCCCAGTAGCTGGGATTACAGGCACCTGCCCCACGCCCGGCTAATTTTTGTAT 70573

Qy       1393 TTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAG 1452
              |||||||||||||| | ||||| |||||||  |||||||| | |||||||||||||| ||||
Db      70574 TTTTAGTAGAGACGGGGTTTCACCATGTTTGCCAGGCTGGTCTTGAACTCCTGACCTCAG 70633

Qy       1453 GTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCC 1512
              ||||||||||  |||||||||||||||  |||||||||||||| ||||| |||| |||| ,
Db      70634 GTGATCCACCTGCCTCAGCCTCCCAAAGTTCTGGGATTACAGGCGTGAGCTACCGCGCCC 70693

Qy       1513 AGCCTG 1518
              ||||||
Db      70694 GGCCTG 70699
```

```
RESULT 31
AP003160
LOCUS       AP003160              130080 bp    DNA     linear   PRI 16-AUG-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 1p36 clone:956M3, complete
            sequence.
ACCESSION   AP003160
VERSION     AP003160.1  GI:15208262
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Machida,T., Ohira,M., Morohashi,A., Mihara,M., Furuta,S., Soeda,E.
            and Nakagawara,A.
  TITLE     Homo sapiens 130,080 genomic DNA of 1p36
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 130080)
  AUTHORS   Machida,T., Ohira,M., Morohashi,A., Mihara,M., Furuta,S., Soeda,E.
            and Nakagawara,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) Akira Nakagawara, Chiba Cancer Center
            Research Institute, Division of Biochemistry; 666-2 Nitona,
            Chuoh-ku, Chiba, Chiba 260-8717, Japan
            (E-mail:akiranak@chiba-cc.pref.chiba.jp, Tel:81-43-264-5431,
            Fax:81-43-265-4459)
COMMENT     Additional author information
            Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A.,
            Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
            The Institute of Physical and Chemical Research (RIKEN) Genomic
            Sciences Center (GSC)
            1-7-22 Suehiro-chou,Tsurumi-ku,Yokohama,Kanagawa, JAPAN zip:
            230-0045
            phone: 81-45-503-9111, fax: 81-45-503-9170
            e-mail: hattori@gsc.riken.go.jp.
FEATURES             Location/Qualifiers
     source          1. .130080
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1p36"
                     /clone="956M3"
ORIGIN

  Query Match             10.6%;  Score 212;  DB 8;  Length 130080;
  Best Local Similarity   78.5%;  Pred. No. 8.3e-38;
  Matches  292;  Conservative    0;  Mismatches   75;  Indels    5;  Gaps    3;

Qy       1192 AGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCA 1251
             |||  |  ||| | ||   | |    || || ||||| |||||||||| |||||||||
Db     103261 AGTTCTCTATTTTTTTAAATTATTCTTTCTTTTTTTTTTTTTTTTTGAGACAGAGTCTTG 103320

Qy       1252 CTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTC 1311
             || |||||||||||||||||| |||||||  || ||| ||||||||||||||||||||||
Db     103321 CTCTGTCACCCAGGCTGGAGTGCAGTGGC-ATGATCTTGGCTCACTGCAACCTCTGCCTC 103379

Qy       1312 CCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC 1369
             |||||||||| ||||||||||  |||||  || || ||| |||||||||| || | |||
Db     103380 CCAGGTTCAAGAGATTCTCCTGTCTCAGCCACCTGAGTAGTTGGGATTACAAGCACGTGC 103439

Qy       1370 --ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAG 1427
                |  ||    |||||||||||||||||||||||||  | ||||| ||||||||| |||
Db     103440 CACCACCCCCAGCTAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAA 103499

Qy       1428 GCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGG 1487
             ||| | ||||||||||||||| |||||||| |||||||||||||||||||||  |||||
Db     103500 GCTGGTCTTGAACTCCTGACCTCAGGTGAACCACCCACCTCAGCCTCCCAAAGTGCTGGG 103559

Qy       1488 ATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAA 1547
             ||||||||| ||||||| | |||||||||   ||| ||  || || ||  | |||||
Db     103560 ATTACAGGCTTGAGCCATCATGCCCAGCCCTATTTTTCCTATTCTTTAAATTTTTCTTTG 103619

Qy       1548 GCCTTCATTTCT 1559
                ||| ||| |
Db     103620 CTTTTCCTTTTT 103631


RESULT 32
HS1056L3/c
LOCUS       HS1056L3              163848 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP5-1056L3 on chromosome 1p35.1-36.13
            Contains a novel gene, the gene for a novel protein (LOC374960), a
            ribosomal protein S14 (RPS14) pseudogene, the NBL1 gene for
            neuroblastoma suppression of tumorigenicity 1, the HTR6 gene for
            5-hydroxytryptamine (serotonin) receptor 6, the 3' end of the gene
            for a novel protein (LOC255104) (FLJ45636) and two CpG islands,
            complete sequence.
ACCESSION   AL031727
VERSION     AL031727.43  GI:21363116
KEYWORDS    HTG; FLJ45636; HTR6; LOC255104; LOC374960; NBL1; RPS14.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 163848)
  AUTHORS   Lad,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 10, 2002 this sequence version replaced gi:9864396.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP5-1056L3 is from the library RPCI-5 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            -------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .163848
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704L031056"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p35.1-36.13"
                     /clone="RP5-1056L3"
                     /clone_lib="RPCI-5"
     misc_feature    1
                     /note="Clone_left_end: RP5-1056L3"
     misc_feature    3341. .3358
                     /note="Sequence from uni-directional primer reads only."
     gene            complement(join(20206. .20289,25114. .25236))
                     /locus_tag="RP5-1056L3.1-001"
     mRNA            complement(join(20206. .20289,25114. .25236))
                     /locus_tag="RP5-1056L3.1-001"
                     /product="putative novel transcript"
                     /note="match: ESTs: Em:AA522834.1 Em:AI732508.1
                     Em:AI791463.1 Em:BX106629.1"
     polyA_site      complement(20206)
                     /locus_tag="RP5-1056L3.1-001"
     polyA_signal    complement(20222. .20227)
                     /locus_tag="RP5-1056L3.1-001"
     gene            join(25399. .25513,42905. .42998,50504. .50551,51878. .51955)
                     /gene="RP5-1056L3.2"
                     /locus_tag="RP5-1056L3.2-003"
     mRNA            join(25399. .25513,42905. .42998,50504. .50551,51878. .51955)
                     /gene="RP5-1056L3.2"
                     /locus_tag="RP5-1056L3.2-003"
                     /product="novel protein"
                     /note="match: ESTs: Em:AA376763.1"
     gene            join(25404. .25513,50504. .50551,51878. .51987,54791. .56714,
                     56994. .58224)
                     /gene="RP5-1056L3.2"
                     /locus_tag="RP5-1056L3.2-001"
     mRNA            join(25404. .25513,50504. .50551,51878. .51987,54791. .56714,
                     56994. .58224)
                     /gene="RP5-1056L3.2"
                     /locus_tag="RP5-1056L3.2-001"
                     /product="novel protein"
                     /note="match: ESTs: Em:AI074989.1 Em:BX107281.1
                     Em:CB852406.1
                     match: cDNAs: Em:AK094318.1"
     gene            join(25405. .25513,45662. .45740,50504. .50551,51878. .51987,
                     54791. .55027)
                     /gene="RP5-1056L3.2"
                     /locus_tag="RP5-1056L3.2-002"
     mRNA            join(25405. .25513,45662. .45740,50504. .50551,51878. .51987,
                     54791. .55027)
                     /gene="RP5-1056L3.2"
                     /locus_tag="RP5-1056L3.2-002"
                     /product="novel protein"
                     /note="match: cDNAs: Em:BC009927.1"
     gene            join(25443. .25513,29191. .29375,50504. .50551,51878. .51987,
```

```
                54791. .55033)
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-004"
     mRNA       join(25443. .25513,29191. .29375,50504. .50551,51878. .51987,
                54791. .55033)
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-004"
                /product="novel protein"
                /note="match: ESTs: Em:BU953571.1"
     gene       join(25445. .25513,50504. .50551,51878. .51987,53796. .54036,
                54791. .54991)
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-005"
     mRNA       join(25445. .25513,50504. .50551,51878. .51987,53796. .54036,
                54791. .54991)
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-005"
                /product="novel protein"
                /note="match: ESTs: Em:CB157536.1 Em:CB162365.1
                Em:CB163300.1"
     CDS        join(25450. .25513,50504. .50551,51878. .51987,54791. .54805)
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-001"
                /standard_name="OTTHUMP00000002794"
                /note="match: proteins: Tr:AAH55791 Tr:Q96G68"
                /codon_start=1
                /product="novel protein"
                /protein_id="CAI19017.1"
                /db_xref="GI:56203284"
                /db_xref="InterPro:IPR007512"
                /db_xref="UniProt/TrEMBL:Q5TGZ0"
                /translation="MSESELGRKWDRCLADAVVKIGTGFGLGIVFSLTFFKRRMWPLA
                FGSGMGLGMAYSNCQHDFQAPYLLHGKYVKEQEQ"
     gene       join(25456. .25513,29191. .29708)
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-006"
     mRNA       join(25456. .25513,29191. .29708)
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-006"
                /product="novel protein"
                /note="match: ESTs: Em:AL602952.1"
     gene       join(25474. .25513,26238. .26257,50504. .50551,51878. .51987,
                54791. .55010)
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-007"
     mRNA       join(25474. .25513,26238. .26257,50504. .50551,51878. .51987,
                54791. .55010)
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-007"
                /product="novel protein"
                /note="match: ESTs: Em:BM504924.1"
     misc_feature 32581. .32758
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-003"
                /note="Sequence confirmed by AP003160 sequenced by RIKEN."
     gene       36518. .36972
                /locus_tag="RP5-1056L3.3-001"
                /pseudo
     CDS        36518. .36972
                /locus_tag="RP5-1056L3.3-001"
                /note="match: proteins: Sw:P06366 Sw:Q08699 Tr:EAA08220
                Tr:Q862B7 Tr:Q90YR0"
                /pseudo
                /codon_start=1
                /product="ribosomal protein S14 (RPS14) pseudogene"
     misc_feature 42637. .42645
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-003"
                /note="1329 bases of Tn10 transposon (J01829) removed
                here. This sequence represents the duplicated flanking
                sequence of the Tn10."
     polyA_signal 55011. .55016
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-004"
     polyA_site 55027
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-004"
     misc_feature 56113. .56126
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-001"
                /note="Sequence confirmed by AP003160 sequenced by RIKEN."
     polyA_signal 58198. .58203
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-001"
     polyA_site 58224
                /gene="RP5-1056L3.2"
                /locus_tag="RP5-1056L3.2-001"
     misc_feature 62566
                /note="Clone_right_end: RP11-296A20"
     misc_feature 65798. .65933
                /note="Sequence confirmed by AP003160 sequenced by RIKEN."
     misc_feature 70460. .70645
                /note="Sequence confirmed by AP003160 sequenced by RIKEN."
     misc_feature 70646. .70783
                /note="Sequence from AP003160 sequenced by RIKEN."

  Query Match            10.6%;  Score 212;  DB 8;  Length 163848;
  Best Local Similarity  78.5%;  Pred. No. 8e-38;
  Matches  292;  Conservative    0;  Mismatches   75;  Indels    5;  Gaps    3;

Qy       1192 AGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCA 1251
              |||  |  ||| | ||   | |    || || ||||| |||||||||| |||||||||
Db      23094 AGTTCTCTATTTTTTTAAATTATTCTTTCTTTTTTTTTTTTTTTTTGAGACAGAGTCTTG 23035

Qy       1252 CTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTC 1311
              || ||||||||||||||||| |||||||  || ||| ||||||||||||||||||||||
Db      23034 CTCTGTCACCCAGGCTGGAGTGCAGTGGC-ATGATCTTGGCTCACTGCAACCTCTGCCTC 22976

Qy       1312 CCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC 1369
              |||||||||| |||||||||  |||||  || || ||| |||||||||| || | |||
Db      22975 CCAGGTTCAAGAGATTCTCCTGTCTCAGCCACCTGAGTAGTTGGGATTACAAGCACGTGC 22916

Qy       1370 --ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAG 1427
                |  ||    |||||||||||||||||||||||  | ||||| ||||||| |||
Db      22915 CACCACCCCCAGCTAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAA 22856

Qy       1428 GCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGG 1487
              ||| | |||||||||||||| |||||||| |||||||||||||||||||||  |||||
Db      22855 GCTGGTCTTGAACTCCTGACCTCAGGTGAACCACCCACCTCAGCCTCCCAAAGTGCTGGG 22796

Qy       1488 ATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAA 1547
              ||||||||| ||||||| | |||||||||   ||| ||  || || ||  | |||||
Db      22795 ATTACAGGCTTGAGCCATCATGCCCAGCCCTATTTTTCCTATTCTTTAAATTTTTCTTTG 22736

Qy       1548 GCCTTCATTTCT 1559
                ||| ||| |
Db      22735 CTTTTCCTTTTT 22724


RESULT 33
AC107299
LOCUS       AC107299               70446 bp    DNA     linear   PRI 09-APR-2002
DEFINITION  Homo sapiens 3 BAC RP11-367K14 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC107299
VERSION     AC107299.3  GI:19697322
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 70446)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
```

Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 70446)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 70446)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 70446)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 70446)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Mar 23, 2002 this sequence version replaced gi:18449819.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
source	1. .70446
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-367K14"
misc_feature	complement(1. .2004)
	/note="overlaps bases 1. .2004 of clone AC072044"
	/function="clone overlap"
repeat_region	246. .408
	/rpt_family="LTR33A"
repeat_region	1054. .1347
	/rpt_family="AluJo"
repeat_region	1348. .1445
	/rpt_family="(TAGA)n"
repeat_region	2650. .2684
	/rpt_family="AT_rich"
repeat_region	complement(3610. .3835)
	/rpt_family="MER46C"
repeat_region	4100. .4339
	/rpt_family="MER46A"
repeat_region	complement(4362. .4519)
	/rpt_family="MIR"
STS	5115. .5240
	/standard_name="73561"
repeat_region	6259. .6338
	/rpt_family="AluJ/FLAM"
repeat_region	8023. .8086
	/rpt_family="AT_rich"
repeat_region	complement(9156. .9359)
	/rpt_family="MER44D"
repeat_region	complement(9346. .9459)
	/rpt_family="MER44D"
repeat_region	11171. .11222
	/rpt_family="(CATATA)n"
repeat_region	complement(16175. .16226)
	/rpt_family="MIR"
repeat_region	complement(16218. .16416)
	/rpt_family="L2"
repeat_region	16424. .19352
	/rpt_family="L1M4"
repeat_region	complement(19379. .19472)

```
                     /rpt_family="AluSc"
     repeat_region   19473. .19833
                     /rpt_family="L1MB8"
     repeat_region   complement(19887. .20219)
                     /rpt_family="L2"
     repeat_region   complement(20237. .20287)
                     /rpt_family="L2"
     repeat_region   20288. .20321
                     /rpt_family="(TTTTA)n"
     repeat_region   complement(20322. .20678)
                     /rpt_family="L2"
     repeat_region   20679. .20951
                     /rpt_family="AluJb"
     repeat_region   20952. .20977
                     /rpt_family="(TAAA)n"
     repeat_region   complement(20978. .21365)
                     /rpt_family="L2"
     repeat_region   21930. .21956
                     /rpt_family="AT_rich"
     repeat_region   23899. .24161
                     /rpt_family="AluSg"
     repeat_region   25101. .25233
                     /rpt_family="MIR"
     repeat_region   25489. .25513
                     /rpt_family="AT_rich"
     repeat_region   27792. .27816
                     /rpt_family="(T)n"
     repeat_region   29516. .29542
                     /rpt_family="(TA)n"
     repeat_region   30921. .31046
                     /rpt_family="MER30"
     repeat_region   31047. .31360
                     /rpt_family="AluY"
     repeat_region   31361. .31461
                     /rpt_family="MER30"
     repeat_region   33737. .33901
                     /rpt_family="MIR"
     repeat_region   35270. .35310
                     /rpt_family="(TGAA)n"

  Query Match           10.6%;  Score 211.8;  DB 8;  Length 70446;
  Best Local Similarity   66.5%;  Pred. No. 1e-37;
  Matches  364;  Conservative    0;  Mismatches  177;  Indels    6;  Gaps    4;

Qy       1203 ATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCC 1262
              ||  ||  ||   | ||||| ||||| |||||||||| |||||||||| || |||  |||
Db      62146 ATAATTCTTCTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCGCTCTGTTGCCC 62205

Qy       1263 AGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAG 1322
              | |||||||||||||| || | |||| ||||||||||||||||| ||| ||||  |||
Db      62206 AAGCTGGAGTACAGTGG-TGCGATCTCAGCTCACTGCAACCTCTGCATCCTAGGTCTAAG 62264

Qy       1323 CGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACC--ATG 1378
              ||||||||||  | ||| |||||| |||||||||||||||| |   |  |  |  ||
Db      62265 CGATTCTCCTGCCTTAGCCTCCCGAGTAGCTGGGATTACAGGCACCCACCACCACTTCTG 62324

Qy       1379 CCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGA 1438
              |||||| |||||||||||||||||||| ||||| |||||||| ||||||| | || | |
Db      62325 GCTAATTATTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCAA 62384

Qy       1439 ACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCAT 1498
              |||||||| ||| ||||||||| |||||||||||||||||||||| ||||||||||||
Db      62385 ACTCCTGACCTCAAGTGATCCCACACACCTCAGCCTCCCAAAGCACTGAGATTACAGGCAT 62444

Qy       1499 GAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGT-TTCTGAAGCCTTCATTT 1557
              |||||||||||||||||| ||||     || || ||  | | || |   |  |  | |
Db      62445 AAGCCACCGTGCCCAGCCTATTTTAAATAATTCTTTAAATATATTTTACGGGGTGTAAAT 62504

Qy       1558 CTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATAT 1617
              ||    ||  ||  |||  |||  |   |  | ||| |    ||    || |  | |
Db      62505 ATAAGACCTCCTTTCTTTCTGAAAAACGTGATTTCAGTGAAAAGTTATTTAACCCTAAAG 62564

Qy       1618 CTTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGAT 1677
                  |  |  |         ||  || |  || |   |||  ||     |  |||
Db      62565 TACCCAAGTAAGCTTAAAAATATTAGATAATTTTAAATTTCTATTGACATCTAGGAAGAT 62624

Qy       1678 GTTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGT 1737
              |   | | |   || ||||| ||    | ||| |   ||||||        | ||| |
Db      62625 ATATAGATACACACACATTATACACACATATATTTACTCATTTATTCATACATTTATTTT 62684

Qy       1738 TATTGAA 1744
              | | |||
Db      62685 TGTAGAA 62691
```

```
RESULT 34
AC117477
LOCUS       AC117477               118135 bp    DNA     linear   PRI 01-JUN-2002
DEFINITION  Homo sapiens 3 BAC RP11-631B21 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC117477
VERSION     AC117477.4  GI:21306535
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 118135)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118135)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2002) Human Genome Sequencing Center, Department
```

```
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 118135)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 118135)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 118135)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 1, 2002 this sequence version replaced gi:20334525.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .118135
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-631B21"
     repeat_region   9. .106
                     /rpt_family="(TAGA)n"
     misc_feature    147. .153
                     /note="1331bp IS2 transposon:ECGALI"
                     /function="bacterial transposon excised"
     repeat_region   1311. .1345
                     /rpt_family="AT_rich"
     repeat_region   complement(2271. .2496)
                     /rpt_family="MER46C"
     repeat_region   2761. .3000
                     /rpt_family="MER46A"
     repeat_region   complement(3023. .3180)
                     /rpt_family="MIR"
     STS             3776. .3901
                     /standard_name="73561"
     repeat_region   4920. .4999
                     /rpt_family="AluJ/FLAM"
     repeat_region   6684. .6747
                     /rpt_family="AT_rich"
     repeat_region   complement(7817. .8020)
                     /rpt_family="MER44D"
     repeat_region   complement(8007. .8120)
                     /rpt_family="MER44D"
     repeat_region   9832. .9883
                     /rpt_family="(CATATA)n"
     repeat_region   complement(14836. .14887)
                     /rpt_family="MIR"
     repeat_region   complement(14879. .15077)
                     /rpt_family="L2"
     repeat_region   15085. .18013
                     /rpt_family="L1M4"
     repeat_region   complement(18040. .18133)
                     /rpt_family="AluSc"
     repeat_region   18134. .18494
                     /rpt_family="L1MB8"
     repeat_region   complement(18548. .18880)
                     /rpt_family="L2"
     repeat_region   complement(18898. .18948)
                     /rpt_family="L2"
     repeat_region   18949. .18982
                     /rpt_family="(TTTTA)n"
     repeat_region   complement(18983. .19339)
                     /rpt_family="L2"
     repeat_region   19340. .19612
                     /rpt_family="AluJb"
     repeat_region   19613. .19638
                     /rpt_family="(TAAA)n"
     repeat_region   complement(19639. .20026)
                     /rpt_family="L2"
     repeat_region   20591. .20617
                     /rpt_family="AT_rich"
     repeat_region   22560. .22822
                     /rpt_family="AluSg"
     repeat_region   23762. .23894
                     /rpt_family="MIR"
     repeat_region   24150. .24174
                     /rpt_family="AT_rich"
     repeat_region   26453. .26477
                     /rpt_family="(T)n"
     repeat_region   28177. .28203
                     /rpt_family="(TA)n"
     repeat_region   29582. .29707
                     /rpt_family="MER30"
     repeat_region   29708. .30021
                     /rpt_family="AluY"
     repeat_region   30022. .30122
                     /rpt_family="MER30"
     repeat_region   32398. .32562
                     /rpt_family="MIR"
     repeat_region   33931. .33971
                     /rpt_family="(TGAA)n"
     repeat_region   complement(34508. .34565)
                     /rpt_family="MIR"
     repeat_region   35991. .36054
                     /rpt_family="GA-rich"

Query Match             10.6%;  Score 211.8;  DB 8;  Length 118135;
Best Local Similarity   66.5%;  Pred. No. 9.3e-38;
Matches  364;  Conservative    0;  Mismatches  177;  Indels    6;  Gaps    4;
```

```
Qy       1203 ATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCC 1262
              ||  ||  ||   | ||||| ||||| |||||||||| |||||||||| || |||  |||
Db      60807 ATAATTCTTCTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCGCTCTGTTGCCC 60866

Qy       1263 AGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAG 1322
              | ||||||||||||||| || | |||| |||||||||||||||||| ||| ||||  |||
Db      60867 AAGCTGGAGTACAGTGG-TGCGATCTCAGCTCACTGCAACCTCTGCATCCTAGGTCTAAG 60925

Qy       1323 CGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACC--ATG 1378
              |||||||||||  | ||| |||||| |||||||||||||||| |   |  |  |   ||
Db      60926 CGATTCTCCTGCCTTAGCCTCCCGAGTAGCTGGGATTACAGGCACCCACCACCACTTCTG 60985

Qy       1379 CCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGA 1438
              |||||| |||||||||||||||||||| ||||| |||||||| ||||||| | || | |
Db      60986 GCTAATTATTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCAA 61045

Qy       1439 ACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCAT 1498
              ||||||||| ||| |||||||||| |||||||||||||||||||||| ||||||||||||
Db      61046 ACTCCTGACCTCAAGTGATCCACACACCTCAGCCTCCCAAAGCACTGAGATTACAGGCAT 61105

Qy       1499 GAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGT-TTCTGAAGCCTTCATTT 1557
              |||||||||||||||||| ||||     || || ||  | | || |   |  |  |  |
Db      61106 AAGCCACCGTGCCCAGCCTATTTTAAATAATTCTTTAAATATATTTTACGGGGTGTAAAT 61165

Qy       1558 CTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATAT 1617
              ||    ||  ||  |||  |||  |     |  | ||| |    ||    || | | | |
Db      61166 ATAAGACCTCCTTTCTTTCTGAAAAACGTGATTTCAGTGAAAAGTTATTTAACCCTAAAG 61225

Qy       1618 CTTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGAT 1677
                 |  |  |        ||   || |  || |    |||   ||     |  |||
Db      61226 TACCCAAGTAAGCTTAAAAATATTAGATAATTTTAAATTTCTATTGACATCTAGGAAGAT 61285

Qy       1678 GTTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGT 1737
              |   | | |   || ||||| ||    | ||| |    ||||||        | ||| |
Db      61286 ATATAGATACACACACATTATACACACATATATTTACTCATTTATTCATACATTTATTTT 61345

Qy       1738 TATTGAA 1744
              | | |||
Db      61346 TGTAGAA 61352
```

```
RESULT 35
AC025415/c
LOCUS       AC025415               181505 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-367K14 map 3, WORKING DRAFT
            SEQUENCE, 23 unordered pieces.
ACCESSION   AC025415
VERSION     AC025415.3  GI:7770520
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 181505)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-367K14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181505)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 181505)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 12, 2000 this sequence version replaced gi:7528375.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8179
                Center clone name: 367_K_14
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 166432 bases at least Q40
                Consensus quality: 174531 bases at least Q30
                Consensus quality: 177652 bases at least Q20
                Insert size: 185000; agarose-fp
                Insert size: 179305; sum-of-contigs
                Quality coverage: 3.5 in Q20 bases; agarose-fp
                Quality coverage: 3.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1123: contig of 1123 bp in length
            *     1124     1223: gap of 100 bp
            *     1224     2181: contig of 958 bp in length
            *     2182     2281: gap of 100 bp
            *     2282     4712: contig of 2431 bp in length
            *     4713     4812: gap of 100 bp
```

```
            *     4813     8198: contig of 3386 bp in length
            *     8199     8298: gap of 100 bp
            *     8299    11409: contig of 3111 bp in length
            *    11410    11509: gap of 100 bp
            *    11510    15116: contig of 3607 bp in length
            *    15117    15216: gap of 100 bp
            *    15217    19171: contig of 3955 bp in length
            *    19172    19271: gap of 100 bp
            *    19272    22792: contig of 3521 bp in length
            *    22793    22892: gap of 100 bp
            *    22893    27631: contig of 4739 bp in length
            *    27632    27731: gap of 100 bp
            *    27732    33828: contig of 6097 bp in length
            *    33829    33928: gap of 100 bp
            *    33929    40616: contig of 6688 bp in length
            *    40617    40716: gap of 100 bp
            *    40717    45477: contig of 4761 bp in length
            *    45478    45577: gap of 100 bp
            *    45578    53703: contig of 8126 bp in length
            *    53704    53803: gap of 100 bp
            *    53804    61654: contig of 7851 bp in length
            *    61655    61754: gap of 100 bp
            *    61755    72106: contig of 10352 bp in length
            *    72107    72206: gap of 100 bp
            *    72207    81246: contig of 9040 bp in length
            *    81247    81346: gap of 100 bp
            *    81347    89996: contig of 8650 bp in length
            *    89997    90096: gap of 100 bp
            *    90097   100217: contig of 10121 bp in length
            *   100218   100317: gap of 100 bp
            *   100318   114540: contig of 14223 bp in length
            *   114541   114640: gap of 100 bp
            *   114641   129350: contig of 14710 bp in length
            *   129351   129450: gap of 100 bp
            *   129451   143367: contig of 13917 bp in length
            *   143368   143467: gap of 100 bp
            *   143468   161284: contig of 17817 bp in length
            *   161285   161384: gap of 100 bp
            *   161385   181505: contig of 20121 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181505
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-367K14"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1123
                     /note="assembly_fragment"
     gap             1124. .1223
                     /estimated_length=100
     misc_feature    1224. .2181
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     gap             2182. .2281
                     /estimated_length=100
     misc_feature    2282. .4712
                     /note="assembly_fragment"
     gap             4713. .4812
                     /estimated_length=100
     misc_feature    4813. .8198
                     /note="assembly_fragment"
     gap             8199. .8298
                     /estimated_length=100
     misc_feature    8299. .11409
                     /note="assembly_fragment"
     gap             11410. .11509
                     /estimated_length=100
     misc_feature    11510. .15116
                     /note="assembly_fragment"
     gap             15117. .15216
                     /estimated_length=100
     misc_feature    15217. .19171
                     /note="assembly_fragment"
     gap             19172. .19271
                     /estimated_length=100
     misc_feature    19272. .22792
                     /note="assembly_fragment"
     gap             22793. .22892
                     /estimated_length=100
     misc_feature    22893. .27631
                     /note="assembly_fragment"
     gap             27632. .27731
                     /estimated_length=100
     misc_feature    27732. .33828
                     /note="assembly_fragment"
     gap             33829. .33928
                     /estimated_length=100
     misc_feature    33929. .40616
                     /note="assembly_fragment"
     gap             40617. .40716
                     /estimated_length=100
     misc_feature    40717. .45477
                     /note="assembly_fragment
                     clone_end:T7

  Query Match           10.6%;  Score 211.8;  DB 14;  Length 181505;
  Best Local Similarity   66.5%;  Pred. No. 8.8e-38;
  Matches  364;  Conservative    0;  Mismatches  177;  Indels    6;  Gaps    4;

Qy       1203 ATGTTTTCTCAGATCTTTTTATTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCC 1262
              ||  ||  ||   | ||||| ||||| ||||||||||||||| |||||| || |||  |||
Db      61043 ATAATTCTTCTTTTTTTTTTTTTTTTTTTTTTTTTGAAACAAAGTCTCGCTCTGTTGCCC 60984

Qy       1263 AGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAG 1322
              | |||||||||||||| || | |||| |||||||||||||||||| ||| ||||  |||
Db      60983 AAGCTGGAGTACAGTGG-TGCGATCTCAGCTCACTGCAACCTCTGCATCCTAGGTCTAAG 60925

Qy       1323 CGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACC--ATG 1378
              ||||||||||  | ||| |||||| ||||||||||||||||| |   |  |  |   ||
Db      60924 CGATTCTCCTGCCTTAGCCTCCCGAGTAGCTGGGATTACAGGCACCCACCACCACTTCTG 60865

Qy       1379 CCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGA 1438
              |||||| |||||||||||||||||||| ||||| |||||||| ||||||| | ||  |
Db      60864 GCTAATTATTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCAA 60805

Qy       1439 ACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCAT 1498
              |||||||| ||| ||||||||| |||||||||||||||||||||| ||||||||||
Db      60804 ACTCCTGACCTCAAGTGATCCACACACCTCAGCCTCCCAAAGCACTGAGATTACAGGCAT 60745

Qy       1499 GAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGT-TTCTGAAGCCTTCATTT 1557
              |||||||||||||||||| ||||     || || ||  | | || |   |  |  |  |
Db      60744 AAGCCACCGTGCCCAGCCTATTTTAAATAATTCTTTAAATATATTTTACGGGGTGTAAAT 60685

Qy       1558 CTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATAT 1617
              ||    ||  ||  |||  |||  |     |  | ||| |    ||    || |  | |
Db      60684 ATAAGACCTCCTTTCTTTCTGAAAAACGTGATTTCAGTGAAAAGTTATTTAACCCTAAAG 60625

Qy       1618 CTTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGAT 1677
                |  |  |          ||  || |  || |    |||   ||      | |||
Db      60624 TACCCAAGTAAGCTTAAAAATATTAGATAATTTTAAATTTCTATTGACATCTAGGAAGAT 60565

Qy       1678 GTTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGT 1737
              |   | | |   || ||||| ||    | ||| |    ||||||        | ||| |
Db      60564 ATATAGATACACACACATTATACACACATATATTTACTCATTTATTCATACATTTATTTT 60505

Qy       1738 TATTGAA 1744
              | | |||
Db      60504 TGTAGAA 60498


RESULT 36
AC161624
LOCUS       AC161624              190264 bp    DNA     linear   HTG 17-MAY-2005
```

```
DEFINITION  Pan troglodytes chromosome UNKNOWN clone CH251-149E7, ***
            SEQUENCING IN PROGRESS ***, 32 unordered pieces.
ACCESSION   AC161624
VERSION     AC161624.1  GI:65819616
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 190264)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Pan troglodytes clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190264)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2005) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            -------------- Project Information -------------------
            Center project name: C_AB0149E07
            ------------------------------------------------------
            ------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 169768 bases at least Q40
            Consensus quality: 175204 bases at least Q30
            Consensus quality: 178674 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 32 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1180: contig of 1180 bp in length
            *     1181     1280: gap of unknown length
            *     1281     2540: contig of 1260 bp in length
            *     2541     2640: gap of unknown length
            *     2641     3939: contig of 1299 bp in length
            *     3940     4039: gap of unknown length
            *     4040     5503: contig of 1464 bp in length
            *     5504     5603: gap of unknown length
            *     5604     7410: contig of 1807 bp in length
            *     7411     7510: gap of unknown length
            *     7511     9136: contig of 1626 bp in length
            *     9137     9236: gap of unknown length
            *     9237    10445: contig of 1209 bp in length
            *    10446    10545: gap of unknown length
            *    10546    12371: contig of 1826 bp in length
            *    12372    12471: gap of unknown length
            *    12472    14015: contig of 1544 bp in length
            *    14016    14115: gap of unknown length
            *    14116    15377: contig of 1262 bp in length
            *    15378    15477: gap of unknown length
            *    15478    17676: contig of 2199 bp in length
            *    17677    17776: gap of unknown length
            *    17777    20317: contig of 2541 bp in length
            *    20318    20417: gap of unknown length
            *    20418    22581: contig of 2164 bp in length
            *    22582    22681: gap of unknown length
            *    22682    26214: contig of 3533 bp in length
            *    26215    26314: gap of unknown length
            *    26315    29199: contig of 2885 bp in length
            *    29200    29299: gap of unknown length
            *    29300    32273: contig of 2974 bp in length
            *    32274    32373: gap of unknown length
            *    32374    37340: contig of 4967 bp in length
            *    37341    37440: gap of unknown length
            *    37441    41969: contig of 4529 bp in length
            *    41970    42069: gap of unknown length
            *    42070    48594: contig of 6525 bp in length
            *    48595    48694: gap of unknown length
            *    48695    55037: contig of 6343 bp in length
            *    55038    55137: gap of unknown length
            *    55138    60618: contig of 5481 bp in length
            *    60619    60718: gap of unknown length
            *    60719    65847: contig of 5129 bp in length
            *    65848    65947: gap of unknown length
            *    65948    71772: contig of 5825 bp in length
            *    71773    71872: gap of unknown length
            *    71873    79030: contig of 7158 bp in length
            *    79031    79130: gap of unknown length
            *    79131    86151: contig of 7021 bp in length
            *    86152    86251: gap of unknown length
            *    86252    97553: contig of 11302 bp in length
            *    97554    97653: gap of unknown length
            *    97654   106778: contig of 9125 bp in length
            *   106779   106878: gap of unknown length
            *   106879   115928: contig of 9050 bp in length
            *   115929   116028: gap of unknown length
            *   116029   124164: contig of 8136 bp in length
            *   124165   124264: gap of unknown length
            *   124265   138817: contig of 14553 bp in length
            *   138818   138917: gap of unknown length
            *   138918   154701: contig of 15784 bp in length
            *   154702   154801: gap of unknown length
            *   154802   190264: contig of 35463 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190264
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="UNKNOWN"
                     /clone="CH251-149E7"
     misc_feature    1. .1180
                     /note="assembly_name:Contig20"
     gap             1181. .1280
                     /estimated_length=unknown
     misc_feature    1281. .2540
                     /note="assembly_name:Contig23"
     gap             2541. .2640
                     /estimated_length=unknown
     misc_feature    2641. .3939
                     /note="assembly_name:Contig24"
     gap             3940. .4039
                     /estimated_length=unknown
     misc_feature    4040. .5503
                     /note="assembly_name:Contig25"
     gap             5504. .5603
                     /estimated_length=unknown
     misc_feature    5604. .7410
                     /note="assembly_name:Contig26"
     gap             7411. .7510
                     /estimated_length=unknown
     misc_feature    7511. .9136
                     /note="assembly_name:Contig27"
     gap             9137. .9236
                     /estimated_length=unknown
     misc_feature    9237. .10445
                     /note="assembly_name:Contig28"
     gap             10446. .10545
                     /estimated_length=unknown
     misc_feature    10546. .12371
                     /note="assembly_name:Contig29"
     gap             12372. .12471
```

```
                /estimated_length=unknown
     misc_feature    12472. .14015
                /note="assembly_name:Contig30"
     gap             14016. .14115
                /estimated_length=unknown
     misc_feature    14116. .15377
                /note="assembly_name:Contig31"
     gap             15378. .15477
                /estimated_length=unknown
     misc_feature    15478. .17676
                /note="assembly_name:Contig32"
     gap             17677. .17776
                /estimated_length=unknown
     misc_feature    17777. .20317
                /note="assembly_name:Contig33"
     gap             20318. .20417
                /estimated_length=unknown
     misc_feature    20418. .22581
                /note="assembly_name:Contig34"
     gap             22582. .22681
                /estimated_length=unknown
     misc_feature    22682. .26214
                /note="assembly_name:Contig35"
     gap             26215. .26314
                /estimated_length=unknown
     misc_feature    26315. .29199
                /note="assembly_name:Contig36"
     gap             29200. .29299
                /estimated_length=unknown
     misc_feature    29300. .32273
                /note="assembly_name:Contig37"
     gap             32274. .32373
                /estimated_length=unknown
     misc_feature    32374. .37340
                /note="assembly_name:Contig38"
     gap             37341. .37440
                /estimated_length=unknown
     misc_feature    37441. .41969
                /note="assembly_name:Contig39"
     gap             41970. .42069
                /estimated_length=unknown
     misc_feature    42070. .48594
                /note="assembly_name:Contig40"
     gap             48595. .48694
                /estimated_length=unknown
     misc_feature    48695. .55037
                /note="assembly_name:Contig41"
     gap             55038. .55137
                /estimated_length=unknown
     misc_feature    55138. .60618
                /note="assembly_name:Contig42"
     gap             60619. .60718
                /estimated_length=unknown

  Query Match             10.6%;  Score 211.6;  DB 14;  Length 190264;
  Best Local Similarity   83.6%;  Pred. No. 9.7e-38;
  Matches  276;  Conservative    0;  Mismatches   49;  Indels    5;  Gaps    3;

Qy       1213 AGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGT 1272
             | || || || ||||| |||||||||| |  || |||  || ||||||||||||||||||
Db     129021 ACATATTCTTCTTTTTTTTTTTTTGAGATGGACTCTTGCTCTGTCACCCAGGCTGGAGT 129080

Qy       1273 ACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCT 1332
              ||| || |    |||||||||||||||| ||| ||||||||||||||||||||||||||
Db     129081 GCAG-GGGTACAATCTCGGCTCACTGCAATCTCCGCCTCCCAGGTTCAAGCGATTCTCCT 129139

Qy       1333 G--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGT 1390
              |  | ||| |||||| |||||||||||||||  |||||  |||||| ||||||||||||||
Db     129140 GCCTTAGCCTCCCGAGTAGCTGGGATTACAG--GCATGTGCCACCACGCCTAATTTTTGT 129197

Qy       1391 ATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTC 1450
              |||||||||||||| | ||||||||||||| ||||||| | || |||||| |||||||
Db     129198 ATTTTTAGTAGAGACGGGGTTTCGCCATGTTGGCCAGGCTGGTCTCGAACTCTTGACTTC 129257

Qy       1451 AGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGC 1510
              |||||||||||||||||||||||||| |||  ||||||||||||||||||||| ||| |
Db     129258 AGGTGATCCACCCACCTCAGCCTCCCGAAGTGCTGGGATTACAGGCATGAGCCAGCGTAC 129317

Qy       1511 CCAGCCTGTTTTCTCAGATCCTGTATTTGT 1540
              ||||||| |||| |    |  | | ||| |
Db     129318 CCAGCCTCTTTTTTTTTTTTTTTTTTTT 129347


RESULT 37
AC008761
LOCUS       AC008761              226170 bp    DNA     linear   PRI 13-JUL-2002
DEFINITION  Homo sapiens chromosome 19 clone CTD-3149D2, complete sequence.
ACCESSION   AC008761
VERSION     AC008761.8  GI:21743747
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 226170)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226170)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 226170)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 226170)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 13, 2002 this sequence version replaced gi:20177645.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.4.
FEATURES             Location/Qualifiers
     source          1. .226170
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTD-3149D2"
ORIGIN

  Query Match             10.6%;  Score 211.6;  DB 8;  Length 226170;
  Best Local Similarity   81.3%;  Pred. No. 9.4e-38;
  Matches  270;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    2;

Qy       1207 TTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGC 1266
              | ||||    | ||||| |||||| |||||||||  ||||||||  || ||||||||||||
Db     184295 TCTCTCTTTTTTTTTTTTTTTTTAATTTTTTTGAGGCAGAGTCTTGCTCTGTCACCCAGGC 184354

Qy       1267 TGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGAT 1326
              |||||| |||||| ||  |||||||||||||||||||| ||||||  ||||||||||||
Db     184355 TGGAGTGCAGTGG-TGCAGTCTCGGCTCACTGCAACCTCCGCCTCCTGGGTTCAAGCGAT 184413

Qy       1327 TCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAAT 1384
              |||||||  ||||| |||| | ||||||||||||||||||| |||  |    | |    |
Db     184414 TCTCCTGCCTCAGCCTCCCCAGTAGCTGGGATTACAGGCGCCTGCCACCATGTCCGGCTT 184473
```

```
Qy       1385 TTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCT 1444
Db     184474 TTTTGTAATTTTAGCAGAGACAGGGTTTCACCATGGTGGCCAGGCTGGTCTTGAACTCCT 184533

Qy       1445 GACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCA 1504
Db     184534 GACCTCAGGTGATCCGCCCGCCTTAGCCTCCCAAAGTGCTGGGATTACAGGCATGGGCCA 184593

Qy       1505 CCGTGCCCAGCCTGTTTTCTCAGATCCTGTAT 1536
Db     184594 CCGTGCCCAGCCCCTCAATTGAGTTTCCGTCT 184625


RESULT 38
D88268/c
LOCUS       D88268                44553 bp    DNA     linear   PRI 26-MAR-2002
DEFINITION  Homo sapiens immunoglobulin lambda gene locus DNA, clone:47H9.
ACCESSION   D88268
VERSION     D88268.1  GI:2114305
KEYWORDS    immunoglobulin light chain.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawasaki,K., Minoshima,S., Nakato,E., Shibuya,K., Shintani,A.,
            Schmeits,J.L., Wang,J. and Shimizu,N.
  TITLE     One-megabase sequence analysis of the human immunoglobulin lambda
            gene locus
  JOURNAL   Genome Res. 7 (3), 250-261 (1997)
   PUBMED   9074928
REFERENCE   2  (bases 1 to 44553)
  AUTHORS   Shimizu,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-1996) Nobuyoshi Shimizu, Keio University, School
            of Medicine; 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan
            (E-mail:shimizu@dmb.med.keio.ac.jp, Tel:03-3351-2370,
            Fax:03-3351-2370)
FEATURES             Location/Qualifiers
     source          1. .44553
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
                     /clone="47H9"
                     /cell_type="human/rodent somatic cell hybrid"
                     /clone_lib="specific cosmid library; LL22NC03"
     misc_feature    1. .44553
                     /note="region V upstream"
                     /db_xref="IMGT/LIGM:D88268"
ORIGIN

  Query Match             10.5%;  Score 211;  DB 8;  Length 44553;
  Best Local Similarity   72.4%;  Pred. No. 1.6e-37;
  Matches  315;  Conservative    0;  Mismatches  115;  Indels    5;  Gaps    3;

Qy       1104 TGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTTAGCCATGTGTGACAAAAGAGAGG 1163
Db      31604 TGTAAAGATAACTCCAATCACAATGTGTTCCTTGTTTTGATCCTGATCTAAACAGACAAT 31545

Qy       1164 CCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTA 1223
Db      31544 TTATGAGATGATTCAAAGTTTGAGCATGGATATTTAATGATATTACAGAAGTATTAATTT 31485

Qy       1224 TTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGT 1283
Db      31484 TTTTTTTTTTTTTTTGAGAAAGAGTCTTGCTCTGTCAGCCAGGCTGGAGTGCAGTGGC-AT 31426

Qy       1284 GGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTC 1341
Db      31425 GATCTCGGCTCACTGCAACCTCTGCCTCCTGGGCTCAAGCAATTCTCCTGCCTCAGCCTC 31366

Qy       1342 CCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTATTTTTAGT 1399
Db      31365 CCGAGTAGCTAGGATTACAGGGGTATGCCACCATGCCCGGCTAATTTTTGTATTTTTAGT 31306

Qy       1400 AGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCC 1459
Db      31305 AGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCC 31246

Qy       1460 ACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGT 1519
Db      31245 ACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGTGCCCGGCCAGA 31186

Qy       1520 TTTCTCAGATCCTGT 1534
Db      31185 AGTATTAAATTTTTT 31171


RESULT 39
AP002893/c
LOCUS       AP002893              178000 bp    DNA     linear   PRI 30-MAR-2002
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-624G17,
            complete sequence.
ACCESSION   AP002893
VERSION     AP002893.3  GI:19879819
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 178000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Apr 1, 2002 this sequence version replaced gi:13469698.
FEATURES             Location/Qualifiers
     source          1. .178000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-624G17"
ORIGIN

  Query Match             10.5%;  Score 211;  DB 8;  Length 178000;
  Best Local Similarity   80.5%;  Pred. No. 1.3e-37;
  Matches  284;  Conservative    0;  Mismatches   65;  Indels    4;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
Db      18298 TCTCTTTCTTCTTTTTTTTTTTGAGATGGAGTCTGGCTCTGTTGCCCAGGCTGGAGTGCA 18239

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
Db      18238 GTGG-TACAATCTCGGCTTACTGCAACCTCTGCCTCCCAGGTTCAAGCATTTCTCCTGCC 18180

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC-CATGCCTAATTTTTGTAT 1392
Db      18179 TCAGCCTCCCGAGTAGCTGGGATTACAGGCGCCCGCACCACACCTGGCTAATTTTTGTAT 18120

Qy       1393 TTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAG 1452
```

```
                  |||||||||||||||| |||||||||||||| ||||||| | || |||||||  || ||||
Db      18119 TTTTAGTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGGTCTAGAACTCCCAACCTCAG 18060

Qy       1453 GTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCC 1512
                  |||||||||||||||||||||||||||| |||||||||||||| |||| ||||||||||
Db      18059 GTGATCCACCCACCTCAGCCTCCCAAAGTACTGGGATTACAGGCGTGAGACACCGTGCCC 18000

Qy       1513 AGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTC 1565
                  |||| |    || |  | ||  ||   || |  |  |     | | | ||| |
Db      17999 AGCCAGGAGGCTTATTTTCTTGATAAATTACCCAGTCTCAGGTATTTCTCTAC 17947


RESULT 40
CR589878/c
LOCUS       CR589878               87071 bp    DNA     linear   PRI 14-SEP-2004
DEFINITION  Human DNA sequence from clone XX-HCC1954_10P22, complete sequence.
ACCESSION   CR589878
VERSION     CR589878.6  GI:51966190
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 87071)
  AUTHORS   Pelan,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2004 this sequence version replaced gi:51950497.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-HCC1954_10P22
            is from a Cancer cell lines BACS library VECTOR: pBACe3.6_BamHI.
FEATURES             Location/Qualifiers
     source          1. .87071
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="XX-HCC1954_10P22"
                     /clone_lib="Cancer cell lines BACS"
ORIGIN

  Query Match             10.5%;  Score 210.8;  DB 8;  Length 87071;
  Best Local Similarity   79.9%;  Pred. No. 1.6e-37;
  Matches  286;  Conservative    0;  Mismatches   67;  Indels    5;  Gaps    3;

Qy       1188 GAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGT 1247
                  || | | | ||  |||| || | |   ||  |||| ||||| |||||||||| |||||||
Db      15944 GACATGAAGTTCCTTATATTATTTTCTATACTTTTTTTTTTTTTTTTTTTGAGACAGAGT 15885

Qy       1248 CTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTG 1307
                  |||||| |||||||||||||||| ||||||| ||  || |||||||||||||||||| |
Db      15884 CTCACTCTGTCACCCAGGCTGGAGCACAGTGG-TGCAGTTTCGGCTCACTGCAACCTCCG 15826

Qy       1308 CCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGC 1365
                  ||||||  |||||||| |||| |||||  ||||| |||||| | |||||||||||||| ||
Db      15825 CCTCCCTAGTTCAAGAGATTTTCCTGCCTCAGCCTCCCGAGTGGCTGGGATTACAGGTGC 15766

Qy       1366 ATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGA 1423
                  | ||  |  ||| |   |||||||||||||||||| ||||||| ||||| |||||||
Db      15765 ACGCCACCACATCCAGGTAATTTTTGTATTTTTAGTTGAGACAGGGTTTCACCATGTTGG 15706

Qy       1424 CCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCAC 1483
                  ||||||| |  |||||||||||||||||| ||||||  |||||||  |||||||||||  |
Db      15705 CCAGGCTGGTTTTGAACTCCTGACTTCAAGTGATCTGCCCACCTTGGCCTCCCAAAGTGC 15646

Qy       1484 TGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
                  |||||||||||||||||||||  |||  | || |||||||   |  | || |  ||
Db      15645 TGGGATTACAGGCATGAGCCACCACGCCTGGGCTATTTTCTCTACTTTTTTAATGTTT 15588


RESULT 41
CR954268/c
LOCUS       CR954268              107537 bp    DNA     linear   PRI 03-MAY-2005
DEFINITION  Human DNA sequence from clone XX-HCC1954_40B13, complete sequence.
ACCESSION   CR954268
VERSION     CR954268.1  GI:63025095
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 107537)
  AUTHORS   Hunter,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-HCC1954_40B13
            is from a Cancer cell lines BACS library VECTOR: pBACe3.6_BamHI.
FEATURES             Location/Qualifiers
     source          1. .107537
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="XX-HCC1954_40B13"
                     /clone_lib="Cancer cell lines BACS"
```

```
ORIGIN

  Query Match             10.5%;  Score 210.8;  DB 8;  Length 107537;
  Best Local Similarity   79.9%;  Pred. No. 1.6e-37;
  Matches  286;  Conservative    0;  Mismatches   67;  Indels    5;  Gaps    3;

Qy       1188 GAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGT 1247
             || | | | ||  |||| || | |    ||  |||| ||||| |||||||||| |||||||
Db      75233 GACATGAAGTTCCTTATATTATTTTCTATACTTTTTTTTTTTTTTTTTTTGAGACAGAGT 75174

Qy       1248 CTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTG 1307
             |||||| |||||||||||||||||| ||||||| ||  || |||||||||||||||||| |
Db      75173 CTCACTCTGTCACCCAGGCTGGAGCACAGTGG-TGCAGTTTCGGCTCACTGCAACCTCCG 75115

Qy       1308 CCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGC 1365
             ||||||  ||||||| |||| |||||  ||||| |||||| | |||||||||||||| ||
Db      75114 CCTCCCTAGTTCAAGAGATTTTCCTGCCTCAGCCTCCCGAGTGGCTGGGATTACAGGTGC 75055

Qy       1366 ATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGA 1423
             | ||  |  ||| |   |||||||||||||||||| |||||| ||||| ||||||||
Db      75054 ACGCCACCACATCCAGGTAATTTTTGTATTTTTAGTTGAGACAGGGTTTCACCATGTTGG 74995

Qy       1424 CCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCAC 1483
             ||||||| |  |||||||||||||||| ||||||  |||||||  ||||||||||  |
Db      74994 CCAGGCTGGTTTTGAACTCCTGACTTCAAGTGATCTGCCCACCTTGGCCTCCCAAAGTGC 74935

Qy       1484 TGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             ||||||||||||||||||||||  |||  | || |||||||   |  | || |  ||
Db      74934 TGGGATTACAGGCATGAGCCACCACGCCTGGGCTATTTTCTCTACTTTTTTAATGTTT 74877


RESULT 42
CR936851
LOCUS       CR936851               138165 bp    DNA     linear   PRI 26-FEB-2005
DEFINITION  Human DNA sequence from clone XX-HCC1954_35H09, complete sequence.
ACCESSION   CR936851
VERSION     CR936851.1  GI:60302467
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 138165)
  AUTHORS   Kay,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-HCC1954_35H09
            is from a Cancer cell lines BACS library VECTOR: pBACe3.6_BamHI.
FEATURES    Location/Qualifiers
     source 1. .138165
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="XX-HCC1954_35H09"
            /clone_lib="Cancer cell lines BACS"
ORIGIN

  Query Match             10.5%;  Score 210.8;  DB 8;  Length 138165;
  Best Local Similarity   79.9%;  Pred. No. 1.5e-37;
  Matches  286;  Conservative    0;  Mismatches   67;  Indels    5;  Gaps    3;

Qy       1188 GAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGT 1247
             || | | | ||  |||| || | |    ||  |||| ||||| |||||||||| |||||||
Db      72260 GACATGAAGTTCCTTATATTATTTTCTATACTTTTTTTTTTTTTTTTTTTGAGACAGAGT 72319

Qy       1248 CTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTG 1307
             |||||| |||||||||||||||||| ||||||| ||  || |||||||||||||||||| |
Db      72320 CTCACTCTGTCACCCAGGCTGGAGCACAGTGG-TGCAGTTTCGGCTCACTGCAACCTCCG 72378

Qy       1308 CCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGC 1365
             ||||||  ||||||| |||| |||||  ||||| |||||| | |||||||||||||| ||
Db      72379 CCTCCCTAGTTCAAGAGATTTTCCTGCCTCAGCCTCCCGAGTGGCTGGGATTACAGGTGC 72438

Qy       1366 ATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGA 1423
             | ||  |  ||| |   |||||||||||||||||| |||||| ||||| ||||||||
Db      72439 ACGCCACCACATCCAGGTAATTTTTGTATTTTTAGTTGAGACAGGGTTTCACCATGTTGG 72498

Qy       1424 CCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCAC 1483
             ||||||| |  |||||||||||||||| ||||||  |||||||  ||||||||||  |
Db      72499 CCAGGCTGGTTTTGAACTCCTGACTTCAAGTGATCTGCCCACCTTGGCCTCCCAAAGTGC 72558

Qy       1484 TGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             ||||||||||||||||||||||  |||  | || |||||||   |  | || |  ||
Db      72559 TGGGATTACAGGCATGAGCCACCACGCCTGGGCTATTTTCTCTACTTTTTTAATGTTT 72616


RESULT 43
CR938745/c
LOCUS       CR938745               164033 bp    DNA     linear   PRI 22-MAR-2005
DEFINITION  Human DNA sequence from clone XX-HCC1954_32J13, complete sequence.
ACCESSION   CR938745
VERSION     CR938745.1  GI:61846497
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 164033)
  AUTHORS   Johnson,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
```

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-HCC1954_32J13 is from a Cancer cell lines BACS library VECTOR: pBACe3.6_BamHI.

```
FEATURES             Location/Qualifiers
     source          1. .164033
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="XX-HCC1954_32J13"
                     /clone_lib="Cancer cell lines BACS"
ORIGIN

  Query Match             10.5%;  Score 210.8;  DB 8;  Length 164033;
  Best Local Similarity   79.9%;  Pred. No. 1.5e-37;
  Matches  286;  Conservative    0;  Mismatches   67;  Indels    5;  Gaps    3;

Qy       1188 GAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGT 1247
             || | | | ||  |||| || | |   ||  |||| ||||| |||||||||| |||||||
Db      50954 GACATGAAGTTCCTTATATTATTTTCTATACTTTTTTTTTTTTTTTTTTTGAGACAGAGT 50895

Qy       1248 CTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTG 1307
             |||||| |||||||||||||||||| ||||||| ||  || |||||||||||||||||| |
Db      50894 CTCACTCTGTCACCCAGGCTGGAGCACAGTGG-TGCAGTTTCGGCTCACTGCAACCTCCG 50836

Qy       1308 CCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGC 1365
             ||||||  ||||||| |||| |||||  ||||| |||||| | |||||||||||||| ||
Db      50835 CCTCCCTAGTTCAAGAGATTTTCCTGCCTCAGCCTCCCGAGTGGCTGGGATTACAGGTGC 50776

Qy       1366 ATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGA 1423
             | ||  |  ||| |   |||||||||||||||||| ||||||| ||||| ||||||||
Db      50775 ACGCCACCACATCCAGGTAATTTTTGTATTTTTAGTTGAGACAGGGTTTCACCATGTTGG 50716

Qy       1424 CCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCAC 1483
             ||||||| |  |||||||||||||||||| ||||||  |||||||  ||||||||||  |
Db      50715 CCAGGCTGGTTTTGAACTCCTGACTTCAAGTGATCTGCCCACCTTGGCCTCCCAAAGTGC 50656

Qy       1484 TGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             |||||||||||||||||||||  |||  | || |||||||    | | || |  ||
Db      50655 TGGGATTACAGGCATGAGCCACCACGCCTGGGCTATTTTCTCTACTTTTTTAATGTTT 50598


RESULT 44
CR936885/c
LOCUS       CR936885              175639 bp    DNA     linear   PRI 01-MAR-2005
DEFINITION  Human DNA sequence from clone XX-HCC1954_35K12, complete sequence.
ACCESSION   CR936885
VERSION     CR936885.1  GI:60417534
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 175639)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-HCC1954_35K12 is from a Cancer cell lines BACS library VECTOR: pBACe3.6_BamHI.

```
FEATURES             Location/Qualifiers
     source          1. .175639
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="XX-HCC1954_35K12"
                     /clone_lib="Cancer cell lines BACS"
ORIGIN

  Query Match             10.5%;  Score 210.8;  DB 8;  Length 175639;
  Best Local Similarity   79.9%;  Pred. No. 1.5e-37;
  Matches  286;  Conservative    0;  Mismatches   67;  Indels    5;  Gaps    3;

Qy       1188 GAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGT 1247
             || | | | ||  |||| || | |   ||  |||| ||||| |||||||||| |||||||
Db      78421 GACATGAAGTTCCTTATATTATTTTCTATACTTTTTTTTTTTTTTTTTTTGAGACAGAGT 78362

Qy       1248 CTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTG 1307
             |||||| |||||||||||||||||| ||||||| ||  || |||||||||||||||||| |
Db      78361 CTCACTCTGTCACCCAGGCTGGAGCACAGTGG-TGCAGTTTCGGCTCACTGCAACCTCCG 78303

Qy       1308 CCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGC 1365
             ||||||  ||||||| |||| |||||  ||||| |||||| | |||||||||||||| ||
Db      78302 CCTCCCTAGTTCAAGAGATTTTCCTGCCTCAGCCTCCCGAGTGGCTGGGATTACAGGTGC 78243

Qy       1366 ATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGA 1423
             | ||  |  ||| |   |||||||||||||||||| ||||||| ||||| ||||||||
Db      78242 ACGCCACCACATCCAGGTAATTTTTGTATTTTTAGTTGAGACAGGGTTTCACCATGTTGG 78183

Qy       1424 CCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCAC 1483
             ||||||| |  |||||||||||||||||| ||||||  |||||||  ||||||||||  |
Db      78182 CCAGGCTGGTTTTGAACTCCTGACTTCAAGTGATCTGCCCACCTTGGCCTCCCAAAGTGC 78123

Qy       1484 TGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             |||||||||||||||||||||  |||  | || |||||||    | | || |  ||
Db      78122 TGGGATTACAGGCATGAGCCACCACGCCTGGGCTATTTTCTCTACTTTTTTAATGTTT 78065


RESULT 45
AC157497/c
LOCUS       AC157497              207664 bp    DNA     linear   PRI 27-APR-2005
DEFINITION  Pan troglodytes BAC clone CH251-490K17 from chromosome unknown,
            complete sequence.
ACCESSION   AC157497
VERSION     AC157497.3  GI:62526284
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 207664)
  AUTHORS   Van Brunt,A., Lewis,S. and Bielicki,L.
```

```
  TITLE     The sequence of Pan troglodytes BAC clone CH251-490K17
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 207664)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2005) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 207664)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2005) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 207664)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2005) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 207664)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2005) Washington University School of Medicine,
            Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO
            63108, USA
COMMENT     On Apr 13, 2005 this sequence version replaced gi:60459078.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: C_AB0490K17


            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The CHORI-251 Chimpanzee BAC library has been constructed at the
            Children's Hospital Oakland Research Institute, BACPAC Resources,
            by Dr. Baoli Zhu.  DNA was isolated from white blood cells obtained
            from a male chimpanzee (Pan troglodytes, 'Clint', Yerkes #C0471;
            birthdate:6-6-80).  The clone and detailed information can be
            obtained from Pieter de Jong and co-workers at
            http://www.bacpac.chori.org.

            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .207664
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="unknown"
                     /clone="CH251-490K17"
                     /clone_lib="CHORI251"
     unsure          28714. .28766
                     /note="Sequence derived from one plasmid subclone."
ORIGIN

  Query Match             10.5%;  Score 210.8;  DB 8;  Length 207664;
  Best Local Similarity   82.9%;  Pred. No. 1.5e-37;

  Matches  277;  Conservative    0;  Mismatches   52;  Indels    5;  Gaps    3;

Qy       1197 TTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTG 1256
             ||  || | | || ||   ||||||  || ||  ||||||| | |||||||||| |||||
Db     132901 TTTCTTTTTTGTTTTCTTTTCTTTTATTTATTTATTTTTTTAACACAGAGTCTCCCTTTG 132842

Qy       1257 TCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 1316
             || |||||||||||||| |||||||    | ||||||||||||||||||||||||||||| ||
Db     132841 TCGCCCAGGCTGGAGTGCAGTGGC-ACGATCTCGGCTCACTGCAACCTCTGCCTCCCGGG 132783

Qy       1317 TTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACC 1372
             |||||| |||||| |||  ||||| ||| || ||||||||||||||||||  ||   |
Db     132782 TTCAAGTGATTCTGCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCGCCCGCCATCG 132723

Qy       1373 ACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTG 1432
               |  | |||||||||||||||||||||||| | ||||| | |||||| ||||||| |
Db     132722 CACCCGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACTATGTTGGCCAGGCTGG 132663

Qy       1433 CCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTAC 1492
              || |||||||||| |||||||||||||||||||  ||||||||||  ||||||||||
Db     132662 TCTGGAACTCCTGACCTCAGGTGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATTAC 132603

Qy       1493 AGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCA 1526
              |||||||||||||| |||| ||||||||||||
Db     132602 AGGCATGAGCCACCGCGCCCGGCCTGTTTTCTCA 132569


Search completed: January 22, 2006, 13:37:59
Job time : 6825.68 secs
```

(older) Number Concepts

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2006, 09:21:22 ; Search time 805.081 Seconds
(without alignments)
16564.866 Million cell updates/sec

Title: US-09-728-552A-3_COPY_19000_21000
Perfect score: 2001
Sequence: 1 aatgtttggatttcaaaggt.....cagggacatgggtgcatagc 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2001	100.0	80595	2	AAV83939	Aav83939 HC-contig
2	1974.2	98.7	80240	2	AAV83940	Aav83940 NC-contig
3	215.8	10.8	60815	11	ACN43882	Acn43882 Human gen
c 4	209.4	10.5	182328	12	ADL08128	Adl08128 Human gen
c 5	208.6	10.4	1425	5	ABA20632	Aba20632 Human ner
c 6	208.2	10.4	409	4	AAK70203	Aak70203 Human imm
7	208	10.4	31813	4	AAL06997	Aal06997 Human rep
8	208	10.4	31813	4	ABA08137	Aba08137 Human ova
9	208	10.4	104245	12	ADG86768	Adg86768 Human clo
10	208	10.4	104245	12	ADL34706	Adl34706 Human PPA
c 11	208	10.4	170245	12	ADP13586	Adp13586 Renal cel
12	206.8	10.3	16424	4	AAK75677	Aak75677 Human imm
c 13	206.8	10.3	16424	4	AAK68448	Aak68448 Human imm
c 14	206.8	10.3	35192	11	ADP75184	Adp75184 Human ADA
c 15	206.8	10.3	118466	13	ABD32861	Abd32861 Human can
16	206.8	10.3	174448	11	ACN43946	Acn43946 Human gen
17	205.6	10.3	48642	12	ADQ97489	Adq97489 Human can
c 18	204.8	10.2	144460	3	AAZ93815	Aaz93815 Olfactory
19	204.8	10.2	168821	11	ACN44262	Acn44262 Human gen
20	203.4	10.2	11343	5	ABA18360	Aba18360 Human ner
21	203.4	10.2	11343	10	ADC21029	Adc21029 Human sec
22	203.4	10.2	11343	10	ABT17025	Abt17025 Human sec
23	203.4	10.2	11343	10	ABZ68170	Abz68170 Human sec
c 24	202.8	10.1	118931	11	ACN44586	Acn44586 Human gen
25	202.6	10.1	124164	12	ADH19244	Adh19244 Human 5-h
26	202.6	10.1	124165	12	ADH19246	Adh19246 Human 5-h
27	202.6	10.1	124167	12	ADF83402	Adf83402 Human 5-h
28	202.6	10.1	124167	12	ADH19240	Adh19240 Human 5-h
29	202.4	10.1	6670	4	AAK70510	Aak70510 Human imm
c 30	202.4	10.1	64829	14	AEA61181	Aea61181 Human FHL
31	202.2	10.1	12149	4	AAS36758	Aas36758 Human car
32	202.2	10.1	12149	10	ADE47452	Ade47452 Human car
33	202.2	10.1	12149	13	ADJ08870	Adj08870 Human car
c 34	202	10.1	160274	14	AEB32377	Aeb32377 Human gen
c 35	202	10.1	160300	14	AEB32388	Aeb32388 Human gen
c 36	201.6	10.1	51256	10	ADB85543	Adb85543 Human sen
c 37	201.6	10.1	51256	10	ADK66161	Adk66161 Human pro
c 38	201.6	10.1	89567	13	ABD33611	Abd33611 Human can
39	201.6	10.1	128034	10	ADE43582	Ade43582 Polymorph
40	201.6	10.1	128034	10	ADE43581	Ade43581 Human IDE
41	201.6	10.1	128034	12	ADH54059	Adh54059 Human IDE
42	201.6	10.1	128034	12	ADH54060	Adh54060 Human IDE
c 43	201.6	10.1	202100	10	ADE43315	Ade43315 Human IDE
c 44	201.6	10.1	202100	12	ADH54357	Adh54357 Human IDE
c 45	201.6	10.1	238417	13	ABD32868	Abd32868 Human can
c 46	201.4	10.1	5881	4	AAL07230	Aal07230 Human rep
c 47	201.4	10.1	5881	4	ABL98778	Abl98778 Human tes
48	201.4	10.1	28066	10	ADC87584	Adc87584 Human GPC
49	201.4	10.1	202100	10	ADE43315	Ade43315 Human IDE
50	201.4	10.1	202100	12	ADH54357	Adh54357 Human IDE
51	201.2	10.1	14209	4	AAK89137	Aak89137 Human dig
52	201.2	10.1	108316	10	ADC87336	Adc87336 Human GPC
c 53	201.2	10.1	110000	10	ADG70184_2	Continuation (3 of
c 54	201.2	10.1	117750	13	ABD32653	Abd32653 Human can
55	201.2	10.1	124987	12	ADP03055	Adp03055 Human hou
56	201.2	10.1	124990	13	ADS88553	Ads88553 Human hou
57	201.2	10.1	124990	13	ADU60195	Adu60195 Housekeep
58	201.2	10.1	154465	6	AAD28763	Aad28763 Human AKA
59	201.2	10.1	158245	6	AAD28762	Aad28762 Human AKA
60	201.2	10.1	161425	4	AAH02340	Aah02340 Human AKA
61	201.2	10.1	162025	4	AAH02339	Aah02339 Human AKA
62	201.2	10.1	162025	6	AAD28758	Aad28758 Human AKA
63	201.2	10.1	162025	6	AAD28759	Aad28759 Human AKA
64	201.2	10.1	162025	13	ADS75958	Ads75958 Human A-k
65	201.2	10.1	162025	13	ADS75959	Ads75959 Human A-k
66	201.2	10.1	162025	13	ADS75963	Ads75963 Human A-k
67	201.2	10.1	162025	13	ADS75962	Ads75962 Human A-k
c 68	201	10.0	1000	2	AAZ27732	Aaz27732 Human DNA
69	201	10.0	11097	10	ADF50893	Adf50893 Promoter
70	201	10.0	27204	11	ACN44866	Acn44866 Human gen
71	201	10.0	81099	11	ACN45018	Acn45018 Human gen
72	201	10.0	94719	10	ADE95902	Ade95902 Human STA
73	201	10.0	94720	9	ADA02654	Ada02654 Human STA
74	201	10.0	94720	10	ADB72392	Adb72392 Human STA
c 75	201	10.0	196686	11	ACN44170	Acn44170 Human gen
76	200.8	10.0	22013	4	AAK85635	Aak85635 Human imm
77	200.8	10.0	22013	5	ABA16084	Aba16084 Human ner
78	200.8	10.0	125515	10	ADL13941	Adl13941 Osteoarth
c 79	200.8	10.0	175077	11	ACN44626	Acn44626 Human gen
80	200.6	10.0	9745	4	AAS36759	Aas36759 Human car
81	200.6	10.0	9745	10	ADE47453	Ade47453 Human car
82	200.6	10.0	9745	13	ADJ08871	Adj08871 Human car
83	200.6	10.0	11301	4	AAK71591	Aak71591 Human imm
c 84	200.6	10.0	110000	10	ACF42745_1	Continuation (2 of
c 85	200.6	10.0	198073	11	ACN44302	Acn44302 Human gen
c 86	200.6	10.0	247509	10	ADL13684	Adl13684 Osteoarth
c 87	200.4	10.0	31970	14	ADZ48481	Adz48481 Human cas
88	200.4	10.0	36359	14	AEB96534	Aeb96534 Human GRA
c 89	200.4	10.0	38504	12	ADQ59185	Adq59185 MSI-H car
c 90	200.4	10.0	44075	9	ADA02774	Ada02774 Human ARH
c 91	200.4	10.0	44075	10	ADB72512	Adb72512 Human ARH
c 92	200.4	10.0	44075	10	ADC85254	Adc85254 Human Arh

c	93	200.4	10.0	44075	12	ADM74369	Adm74369 Human car
c	94	200.4	10.0	47319	4	AAK72230	Aak72230 Human imm
c	95	200.4	10.0	47319	4	AAK64813	Aak64813 Human imm
c	96	200.4	10.0	110000	13	ABD32780_3	Continuation (4 of
	97	200.4	10.0	112132	6	ABK90888	Abk90888 Human ATP
	98	200.4	10.0	112132	11	ADM56267	Adm56267 Human ATP
	99	200.4	10.0	112132	14	ADV85874	Adv85874 Human pro
	100	200.4	10.0	171158	12	ADQ97894	Adq97894 Human can

ALIGNMENTS

RESULT 1
AAV83939
ID AAV83939 standard; DNA; 80595 BP.
XX
AC AAV83939;
XX
DT 03-MAR-1999 (first entry)
XX
DE HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 8; Fig 6; 540pp; English.
XX
CC The present sequence represents the HC-contig derived from normal human
CC chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC produce an unusual chromosomal marker referred to as mardel(10). The
CC mardel(10) marker is mitotically stable and contains a functional
CC neocentromere at a location regarded as non-centromeric. This
CC neocentromere maps to q25.2 on chromosome 10. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in a
CC compatible cell, of replicating, acting as extra-chromosomal element and
CC segregating during cell division. The sequences can be used to construct
CC artificial chromosomes for use in gene therapy comprising a replicable,
CC segregating nucleic acid that confers a specific phenotype on cells.
CC Human artificial chromosomes can propagate in human cells and carry large
CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC they are not mutagenic. The artificial chromosomes are also useful for
CC generation of transgenic plants and animals, in production of proteins
CC and to make diagnostic reagents, e.g. for expression of cytokines,
CC receptors and growth factors, or to increase the copy number of a gene in
CC a cell. The constructs may also be used for functional and structural
CC analysis of chromosomes
XX

SQ Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

Query Match 100.0%; Score 2001; DB 2; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19000 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 19059

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19060 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 19119

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19120 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 19179

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19180 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 19239

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19240 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 19299

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19300 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19359

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19360 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19419

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19420 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19479

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19480 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19539

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19540 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19599

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19600 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19659

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19660 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19719

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19720 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19779

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTTTCAGGTGGGATGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19780 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTTTCAGGTGGGATGG 19839

Qy        841 TAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGCT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19840 TAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGCT 19899

Qy        901 CTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19900 CTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCCC 19959

Qy        961 GTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19960 GTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTAG 20019
```

```
Qy       1021 AATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTAT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20020 AATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTAT 20079

Qy       1081 AGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20080 AGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTT 20139

Qy       1141 AGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20140 AGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGA 20199

Qy       1201 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCAC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20200 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCAC 20259

Qy       1261 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20260 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 20319

Qy       1321 AGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20320 AGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCC 20379

Qy       1381 TAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAAC 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20380 TAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAAC 20439

Qy       1441 TCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20440 TCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGA 20499

Qy       1501 GCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTA 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20500 GCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTA 20559

Qy       1561 TCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20560 TCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTT 20619

Qy       1621 TGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTT 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20620 TGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTT 20679

Qy       1681 ATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTAT 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20680 ATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTAT 20739

Qy       1741 TGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTAAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20740 TGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTAAG 20799

Qy       1801 TGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTGAC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20800 TGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTGAC 20859

Qy       1861 TGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACATGC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20860 TGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACATGC 20919

Qy       1921 TGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACGCA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20920 TGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACGCA 20979

Qy       1981 ACAGGGACATGGGTGCATAGC 2001
             |||||||||||||||||||||
Db      20980 ACAGGGACATGGGTGCATAGC 21000
```

RESULT 2
AAV83940
ID AAV83940 standard; DNA; 80240 BP.

XX
AC AAV83940;
XX
DT 03-MAR-1999 (first entry)
XX
DE NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 9; Fig 16A; 540pp; English.
XX
CC The present sequence represents the NC-contig derived from a mutated
CC human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC mitotically stable and contains a functional neocentromere at a location
CC regarded as non-centromeric. This neocentromere maps to q25.2 on
CC chromosome 10. The specification describes nucleic acid sequences derived
CC from a eukaryotic chromosome, including a neocentromere or its functional
CC derivative or hybrid, that are able, in a compatible cell, of
CC replicating, acting as extra-chromosomal element and segregating during
CC cell division. The sequences can be used to construct artificial
CC chromosomes for use in gene therapy comprising a replicable, segregating
CC nucleic acid that confers a specific phenotype on cells. Human artificial
CC chromosomes can propagate in human cells and carry large amounts of DNA
CC (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC mutagenic. The artificial chromosomes are also useful for generation of
CC transgenic plants and animals, in production of proteins and to make
CC diagnostic reagents, e.g. for expression of cytokines, receptors and
CC growth factors, or to increase the copy number of a gene in a cell. The
CC constructs may also be used for functional and structural analysis of
CC chromosomes
XX
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

Query Match 98.7%; Score 1974.2; DB 2; Length 80240;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1998; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```
Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18736 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 18795

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18796 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 18855

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      18856 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 18915

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18916 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 18975

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18976 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 19035

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      19036 CATATGCAGGAGTGTCACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19095

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19096 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19155

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19156 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19215

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19216 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19275

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19276 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19335

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19336 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19395

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19396 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19455

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19456 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19515

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTT-TCAGGTGGGATG 839
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      19516 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTCTCAGGTGGGATG 19575

Qy        840 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19576 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 19635

Qy        900 TCTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCC 959
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      19636 TCTGACTGGGGACCCTAAATATGCAAAAGTTAATAGGCTCTTCATGCAGAATATGAACCC 19695

Qy        960 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19696 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 19755

Qy       1020 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19756 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 19815

Qy       1080 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19816 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 19875

Qy       1140 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19876 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 19935

Qy       1200 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCACTTTGTCA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19936 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCACTTTGTCA 19995
```

```
Qy       1260 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19996 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 20055

Qy       1320 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20056 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 20115

Qy       1380 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20116 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 20175

Qy       1440 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20176 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 20235

Qy       1500 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA-TTTGTTTCTGAAGCCTTCATTTC 1558
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      20236 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTTGTTTCTGAAGCCTTCATTTC 20295

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20296 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 20355

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20356 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 20415

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20416 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 20475

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20476 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 20535

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20536 ATTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 20595

Qy       1859 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 1918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20596 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 20655

Qy       1919 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 1978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20656 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 20715

Qy       1979 CAACAGGGACATGGGTGCATAGC 2001
              |||||||||||||||||||||||
Db      20716 CAACAGGGACATGGGTGCATAGC 20738
```

```
RESULT 3
ACN43882
ID   ACN43882 standard; DNA; 60815 BP.
XX
AC   ACN43882;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG1777139.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
```

```
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 52; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 60815 BP; 14386 A; 15367 C; 17015 G; 13945 T; 0 U; 102 Other;

  Query Match             10.8%;  Score 215.8;  DB 11;  Length 60815;
  Best Local Similarity   71.4%;  Pred. No. 7e-44;
  Matches  312;  Conservative    0;  Mismatches  122;  Indels    3;  Gaps    2;

Qy       1102 CATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTTAGCCATGTGTGACAAAAGAGA 1161
             | |||     ||| ||  |  |       |  ||||| | ||||   |||    |
Db      36062 CCTGATGGACCTTGAGTTCAGATCCAGGTGCCATTTTGTGGCCAAAAGTGTACTACCCTT 36121

Qy       1162 GGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTT 1221
               | |   |||| |    |  | |   |  | ||  |     | ||||   | ||||
Db      36122 TCCAAGCGCAGTCTCCCCACCCTCAATTTAGTTTCTTTCTCTCTCTCTCTCTCTTTTTTT 36181

Qy       1222 TATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCT 1281
             | |||||| |||||||||  ||||||||  || ||||||||||||||||||| |||||| |
Db      36182 TTTTTTTAATTTTTTTGAGGCAGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGTGG-T 36240

Qy       1282 GTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCT 1339
             |  |||||||||||||||||||| ||||||  |||||||||||||||||||  |||||
Db      36241 GCAGTCTCGGCTCACTGCAACCTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCC 36300

Qy       1340 TCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATTTTTAGT 1399
              |||||| |||||||||||||||||| |||  |     | |    |||||||| ||||||
Db      36301 TCCCGAGTAGCTGGGATTACAGGCGCCTGCCACCATGTCCGGCTTTTTTGTAATTTTAGC 36360

Qy       1400 AGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCC 1459
              |||||||| ||||| ||||| || ||||||| | |||||||||||||| ||||||||||
Db      36361 AGAGACAGGGTTTCACCATGGTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCC 36420

Qy       1460 ACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGT 1519
               ||| ||| |||||||||||||  |||||||||||||||| |||||||||||||||  |
Db      36421 GCCCGCCTTAGCCTCCCAAAGTGCTGGGATTACAGGCATGGGCCACCGTGCCCAGCCCCT 36480

Qy       1520 TTTCTCAGATCCTGTAT 1536
                | || | | || |
Db      36481 CAATTGAGTTTCCGTCT 36497


RESULT 4
ADL08128/c
ID   ADL08128 standard; DNA; 182328 BP.
XX
AC   ADL08128;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human gene associated with low HDL-C PRCP.
XX
KW   Human; ds; SNP; single nucleotide polymorphism;
KW   high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;
KW   coronary artery disease; gene.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(63956,C)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   US2004043389-A1.
XX
PD   04-MAR-2004.
XX
PF   04-SEP-2002; 2002US-00235192.
XX
PR   04-SEP-2002; 2002US-00235192.
XX
PA   (VITI-) VITIVITY INC.
XX
PI   Mccarthy J;
XX
DR   WPI; 2004-214170/20.
XX
PT   Determining whether a subject has, or is at risk of developing, an
PT   abnormally low high density lipoprotein-C (HDL-C) level comprises
PT   detecting an allelic variant of a polymorphic region from any of a set of
PT   27 genes.
XX
PS   Disclosure; SEQ ID NO 47; 37pp; English.
XX
CC   The invention relates to determining whether a subject has, or is at risk
CC   of developing, an abnormally low high density lipoprotein-C (HDL-C) level
CC   comprises determining whether the subject has an allelic variant of a
CC   polymorphic region from any of 27 genes (alleles listed in Table 5 of the
CC   specification). Also included are determining whether a male subject has,
CC   or is at risk of developing, an abnormally low HDL-C level, comprising
CC   determining whether the male subject has an allelic variant of a
CC   polymorphic region listed in Table 5 which is associated with abnormally
CC   low HDL-C levels in males, and determining whether a female subject has,
CC   or is at risk of developing, an abnormally low HDL-C level, comprising
CC   determining whether the female subject has an allelic variant of a
CC   polymorphic region listed in Table 5 which is associated with abnormally
CC   low HDL-C levels in females. The allelic variant in determining whether a
CC   subject has, or is at risk of developing, an abnormally low HDL-C level
CC   is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBE1
CC   1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The
CC   allelic variant in determining whether a male subject has, or is at risk
CC   of developing, an abnormally low HDL-C level, LRP1 3 CC or CT, PAI2 4 GG,
CC   or PPARG 1 CG, or their complements. The allelic variants are also COL5A2
CC   1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their
CC   complements. The methods are useful for diagnosing (a predisposition to)
CC   abnormally low levels of low high density lipoprotein-C (HDL-C) in a
CC   subject, The methods are useful in diagnosing (a predisposition to) or
CC   prognosticating diseases and disorders associated with abnormal lipid
CC   levels such as vascular and metabolic diseases, e.g., coronary artery
CC   disease. The present sequence is a human gene containing a SNP (single
CC   nucleotide polymorphism associated with low high density lipoprotein-C
CC   (HDL-C) levels.
XX
SQ   Sequence 182328 BP; 56638 A; 36576 C; 34943 G; 54171 T; 0 U; 0 Other;

  Query Match             10.5%;  Score 209.4;  DB 12;  Length 182328;
  Best Local Similarity   78.2%;  Pred. No. 4.8e-42;
  Matches  290;  Conservative    0;  Mismatches   76;  Indels    5;  Gaps    3;

Qy       1176 TGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTT 1235
```

```
                || | ||||         | ||   || | | ||      | | ||| ||||| ||||||
Db      83684 TGCATTTGTGTTTTTTTCTTTTCTTTTCTTTCTTTCTTTCTTTTTTTTTTTTTTTT 83625

Qy       1236 TTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCA 1295
              |||| || ||| ||| || |||| |||||||||||||| || ||||  || ||||||||||
Db      83624 TTGAGACGGAGCCTCGCTCTGTCTCCCAGGCTGGAGTGCAATGGC-ATGATCTCGGCTCA 83566

Qy       1296 CTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGG 1353
              |||||||||||||||||| ||||||||||||||||||  | ||| |||| | |||||||
Db      83565 CTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTTAGCCTCCCAAGTAGCTGG 83506

Qy       1354 GATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTC 1413
              ||| ||||  |||  ||||||  ||||| |||||||||||||||||||||| | |||||
Db      83505 GATCACAG--GCACCCACCACTGTGCCTGATTTTTGTATTTTTAGTAGAGACGGGGTTTC 83448

Qy       1414 GCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCT 1473
              |||||||| ||||||| |  | ||||||||||  |||||||||||||| |||||||||
Db      83447 ACCATGTTGGCCAGGCTGGTTTCGAACTCCTGATCTCAGGTGATCCACCCGCCTCAGCCT 83388

Qy       1474 CCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTG 1533
              |||||||  |||||||||||||| |||||||||| |||| |||| |||  |    || |
Db      83387 CCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCCTCTTTCTTTCTTTCTTT 83328

Qy       1534 TATTTGTTTCT 1544
              ||||| ||| |
Db      83327 TATTTATTTTT 83317
```

RESULT 5
ABA20632/c
ID ABA20632 standard; DNA; 1425 BP.
XX
AC ABA20632;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 12963.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.

```
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250391P.
PR   01-DEC-2000; 2000US-0251160P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system cancers
PT   and metastases.
XX
PS   Disclosure; SEQ ID NO 12963; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1425 BP; 368 A; 344 C; 344 G; 369 T; 0 U; 0 Other;

  Query Match             10.4%;  Score 208.6;  DB 5;  Length 1425;
  Best Local Similarity   82.6%;  Pred. No. 1e-42;
  Matches  275;  Conservative    0;  Mismatches   54;  Indels    4;  Gaps    3;

Qy       1220 TTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGG 1279
             ||| ||||| |||||||| | ||  ||||||||| |||||||||| |||||||||| ||
Db       1302 TTTTTTTTTTTTTTTTTTAGACTAAGTCTCACTCTGTCACCCAGACTGGAGTACAGAGG 1243

Qy       1280 CTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAG 1337
                |||||||||||||||||||| ||||||||||||||||||||||||||||  ||||
Db       1242 AACAA-TCTCGGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGCGATTCTCCTGCCTCAG 1184

Qy       1338 CTTCCCGAATAGCTGGGATTACAGGCGCATG-CACCACCATGCCTAATTTTTGTATTTTT 1396
              | | || | |||||||||||||||||  | | |||   |  | ||||||||||||||||
Db       1183 CCTTCCTAGTAGCTGGGATTACAGGCATAAGCCACATGCCCGGCTAATTTTTGTATTTTT 1124

Qy       1397 AGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGA 1456
              |||||||||||||||| |||| ||| ||||||| | || |||||||||| ||| ||||
Db       1123 AGTAGAGACAGAGTTTCACCATATTGGCCAGGCTGGTCTCGAACTCCTGACCTCAAGTGA 1064

Qy       1457 TCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCC 1516
              |||||| |||||||||||||||| |||||||| ||||||||||||||||||||||||||
Db       1063 TCCACCAACCTCAGCCTCCCAAAGCTCTGGGATTATAGGCATGAGCCACCGTGCCCAGCC 1004

Qy       1517 TGTTTTCTCAGATCCTGTATTTGTTTCTGAAGC 1549
                 ||| ||||||  |  |   | || ||  ||
Db       1003 GAATTTTTCAGATTTTAGAAAGGATTATGCTGC 971


RESULT 6
AAK70203/c
ID   AAK70203 standard; DNA; 409 BP.
XX
AC   AAK70203;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25015.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
```

PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 25015; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention

```
XX
SQ   Sequence 409 BP; 132 A; 78 C; 125 G; 74 T; 0 U; 0 Other;

  Query Match            10.4%;  Score 208.2;  DB 4;  Length 409;
  Best Local Similarity  82.9%;  Pred. No. 7.7e-43;
  Matches  286;  Conservative    0;  Mismatches   53;  Indels    6;  Gaps    4;

Qy       1201 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCAC 1260
             || | |||| |     ||||| | |||  |||||||| |||||||||| || | |||
Db        407 TTCTTTTTTTTTCTCCTTTTTTTCTTTTCTTTTTTTGAGACAGAGTCTCGCTGTTTCAT 348

Qy       1261 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 1320
             |||||||||||| |||||| |||| |||||||||||||||||||| ||||||||||||||
Db        347 CCAGGCTGGAGTGCAGTGG-TGTGATCTCGGCTCACTGCAACCTC-GCCTCCCAGGTTCA 290

Qy       1321 AGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA-- 1376
             ||| |||||||||  |||||||||| | ||||||||||||||||||||  |  ||
Db        289 AGCAATTCTCCTGCCTCAGCTTCCCAAGTAGCTGGGATTACAGGCGCATGCCACCACACC 230

Qy       1377 TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTT 1436
              | |||||||||||||||||||||||||| ||||| |||||||| ||||||| | |||
Db        229 CGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTT 170

Qy       1437 GAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGC 1496
             |||||||||| |||||| |||| |||||||| |||||||||  ||||||||||||||
Db        169 GAACTCCTGACCTCAGGTAATCCGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGC 110

Qy       1497 ATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
              | ||| |||| |||||||||| | || ||  ||  |  | || ||
Db        109 AAGAGACACCATGCCCAGCCTCTCTTTTCTTATAATCAACTTCTT 65


RESULT 7
AAL06997
ID   AAL06997 standard; DNA; 31813 BP.
XX
AC   AAL06997;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 9685.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
```

PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 9685; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 31813 BP; 7330 A; 7632 C; 7753 G; 9098 T; 0 U; 0 Other;

Query Match 10.4%; Score 208; DB 4; Length 31813;
Best Local Similarity 78.0%; Pred. No. 5.3e-42;
Matches 276; Conservative 0; Mismatches 75; Indels 3; Gaps 2;

```
Qy      1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| ||||||| || |  ||||||  || ||||||||||||||||||| ||
Db     21355 TTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCA 21414

Qy      1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| |||  |||| |||||  |||||||||||||||||||||||||| ||||||||||
```

```
Db     21415 GTGG-TGTAATCTCAGCTCAGCGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGCC 21473

Qy      1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATT 1393
             ||||| |||| | |||||||||| ||||||| | |||  |  || ||      ||||||||
Db     21474 TCAGCCTCCCAAGTAGCTGGGACTACAGGCACGTGCCACCACACGCGGCTAATTTGTATT 21533

Qy      1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             |||||||||||||| |||||||||||||| ||||||| | |||||||||||||| |||||
Db     21534 TTTAGTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGG 21593

Qy      1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||||||||||| ||||||||||  |||||||||||||||||||||| |  |||
Db     21594 TGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCTCCCG 21653

Qy      1514 GCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTT 1567
             ||| |  | |    | |    |  || |   || |  ||| | |||| ||
Db     21654 GCCCGGAATTTTCTTTATTTCTATACATTTTACTTCCGTTGTTTTTTTCTTTTT 21707
```

RESULT 8
ABA08137
ID ABA08137 standard; DNA; 31813 BP.
XX
AC ABA08137;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 932.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ds.
XX
OS Homo sapiens.
XX
PN WO200155325-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001345.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488786/53.
XX
PT New isolated ovarian and/or breast cancer related nucleic acids and
PT polypeptides, useful for diagnosing, treating and/or preventing human
PT diseases and disorders, particularly ovarian and/or breast cancer.
XX
PS Disclosure; SEQ ID NO 932; 577pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA07454-ABA08224) and proteins
CC (ABB10743-ABB10980) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 31813 BP; 7330 A; 7632 C; 7753 G; 9098 T; 0 U; 0 Other;

Query Match 10.4%; Score 208; DB 4; Length 31813;
Best Local Similarity 78.0%; Pred. No. 5.3e-42;
Matches 276; Conservative 0; Mismatches 75; Indels 3; Gaps 2;

Qy 1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
Db 21355 TTTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCA 21414

```
Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| |||  |||| |||||  |||||||||||||||||||||||||| ||||||||||
Db      21415 GTGG-TGTAATCTCAGCTCAGCGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGCC 21473

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATT 1393
             ||||| |||| | ||||||||| ||||||| | |||  |  || ||      ||||||||
Db      21474 TCAGCCTCCCAAGTAGCTGGGACTACAGGCACGTGCCACCACACGCGGCTAATTTGTATT 21533

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             |||||||||||||| |||||||||||||| |||||||| | ||||||||||||| |||||
Db      21534 TTTAGTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGG 21593

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||||||||||| |||||||||| ||||||||||||||||||||||| |  |||
Db      21594 TGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCTCCCG 21653

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTT 1567
             ||| |  | |   |  |   |  || |   || |  ||| | |||| ||
Db      21654 GCCCGGAATTTTCTTTATTTCTATACATTTTACTTCCGTTGTTTTTTTCTTTTT 21707
```

RESULT 9
ADG86768
ID ADG86768 standard; DNA; 104245 BP.
XX
AC ADG86768;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human clone RP1-109F14 from chromosome 6p21.2-21.3.
XX
KW Human; ds; PPAR delta; peroxisome proliferative activated receptor delta;
KW antisense gene therapy; cytostatic; osteopathic; antidiabetic; cancer;
KW osteoporosis; diabetes; endocrine disorder.
XX
OS Homo sapiens.
XX
PN US2003224514-A1.
XX
PD 04-DEC-2003.
XX
PF 31-MAY-2002; 2002US-00160807.
XX
PR 31-MAY-2002; 2002US-00160807.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Gaarde W, Freier SM, Watt AT;
XX
DR WPI; 2004-022078/02.
XX
PT New antisense oligonucleotides of 8-80 nucleobases, useful for treating
PT cancer, diabetes, osteoporosis or various endocrine disorders.
XX
PS Example 13; SEQ ID NO 4; 155pp; English.
XX
CC The invention relates to an antisense oligonucleotide comprising 8-80
CC nucleobases in length targeted to the coding region of a nucleic acid
CC molecule encoding PPAR-delta (peroxisome proliferative activated receptor
CC delta), where the antisense compound inhibits the expression of the PPAR-
CC delta and has any of the 66 sequences of 20 amino acids fully defined in
CC the specification. Also included are a compound of 8-80 nucleobases in
CC length that specifically hybridises with at least an 8-nucleobase portion
CC of a preferred target region on a nucleic acid molecule encoding PPAR-
CC delta and a composition comprising the antisense oligonucleotide and a
CC carrier. The antisense oligonucleotide comprises at least one modified
CC internucleoside linkage (preferably a phosphorothioate linkage), at least
CC one sugar moiety (preferably 2'-O-methoxyethyl moiety) and at least one
CC modified nucleobase (which is a 5-methyl cytosine). The antisense
CC compounds are useful for treating cancer, osteoporosis, diabetes or
CC various endocrine disorders. The Human PPAR delta gene is located on
CC chromosome 6p21. The present sequence is a PPAR delta genomic sequence
CC used as a target for the antisense oligonucleotides of the invention.
XX
SQ Sequence 104245 BP; 25340 A; 24979 C; 26046 G; 27880 T; 0 U; 0 Other;

  Query Match   10.4%; Score 208; DB 12; Length 104245;
  Best Local Similarity 78.0%; Pred. No. 8.6e-42;
  Matches 276; Conservative   0; Mismatches   75; Indels   3; Gaps   2;

```
Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| ||||||| || |  ||||||  || ||||||||||||||||||| ||
Db      49396 TTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCA 49455

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| |||  |||| |||||  |||||||||||||||||||||||||| ||||||||||
Db      49456 GTGG-TGTAATCTCAGCTCAGCGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGCC 49514

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATT 1393
             ||||| |||| | ||||||||| ||||||| | |||  |  || ||      ||||||||
Db      49515 TCAGCCTCCCAAGTAGCTGGGACTACAGGCACGTGCCACCACACGCGGCTAATTTGTATT 49574

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             |||||||||||||| |||||||||||||| |||||||| | ||||||||||||| |||||
Db      49575 TTTAGTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGG 49634

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||||||||||| |||||||||| ||||||||||||||||||||||| |  |||
Db      49635 TGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCTCCCG 49694

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTT 1567
             ||| |  | |   |  |   |  || |   || |  ||| | |||| ||
Db      49695 GCCCGGAATTTTCTTTATTTCTATACATTTTACTTCCGTTGTTTTTTTCTTTTT 49748
```

RESULT 10
ADL34706
ID ADL34706 standard; DNA; 104245 BP.
XX
AC ADL34706;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human PPAR-delta DNA fragment SEQ ID 4.
XX
KW antisense; PPAR-delta; human; hybridisation; inhibitor;
KW phosphorothioate linkage; 2'-O-methoxyethyl sugar; 5-methylcytosine;
KW hyperproliferative disorder; cancer; cytostatic; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN US2004063129-A1.
XX
PD 01-APR-2004.
XX
PF 05-SEP-2003; 2003US-00655847.
XX
PR 31-MAY-2002; 2002US-00160807.
XX
PA (GAAR/) GAARDE W.
PA (FREI/) FREIER S M.
PA (WATT/) WATT A T.
XX
PI Gaarde W, Freier SM, Watt AT;
XX
DR WPI; 2004-282460/26.
XX
PT New antisense oligonucleotide, having a sequence targeted to a nucleic
PT acid encoding PPAR-delta, useful for preparing a composition for treating
PT hyperproliferative disorder, e.g., cancer.
XX
PS Example 15; SEQ ID NO 4; 0pp; English.
XX
CC This invention describes novel antisense oligonucleotides targeted to a
CC nucleic acid encoding PPAR-delta, which specifically hybridise to and

```
CC   inhibit expression of PPAR-delta. The oligonucleotide specifically
CC   hybridises with at least an 8-nucleobase portion of an active site on the
CC   nucleic acid molecule encoding the PPAR-delta and comprises at least one
CC   modified internucleoside linkage, which is a phosphorothioate linkage, at
CC   least one modified sugar moiety, which is a 2'-O-methoxyethyl sugar
CC   moiety or at least one modified nucleobase, which is a 5-methylcytosine.
CC   The antisense oligonucleotides are useful for preparing a composition for
CC   treating hyperproliferative disorders, e.g., cancer. The oligonucleotides
CC   of the invention have cytostatic activity and can be used for gene
CC   therapy.
XX
SQ   Sequence 104245 BP; 25340 A; 24979 C; 26046 G; 27880 T; 0 U; 0 Other;

  Query Match             10.4%;  Score 208;  DB 12;  Length 104245;
  Best Local Similarity   78.0%;  Pred. No. 8.6e-42;
  Matches  276;  Conservative    0;  Mismatches   75;  Indels    3;  Gaps    2;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| ||||||| || |  ||||||  || ||||||||||||||||||| ||
Db      49396 TTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCA 49455

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| |||  |||| |||||  |||||||||||||||||||||||| ||||||||||
Db      49456 GTGG-TGTAATCTCAGCTCAGCGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGCC 49514

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATT 1393
             ||||| |||| | ||||||||| ||||||| | |||  |  || ||      ||||||||
Db      49515 TCAGCCTCCCAAGTAGCTGGGACTACAGGCACGTGCCACCACACGCGGCTAATTTGTATT 49574

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             |||||||||||||| |||||||||||||| ||||||| | |||||||||||||| |||||
Db      49575 TTTAGTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGG 49634

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||||||||||| ||||||||||  |||||||||||||||||||||| |  |||
Db      49635 TGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCTCCCG 49694

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTT 1567
             ||| |   | |    |  |    |   || |    || |  ||| | |||| ||
Db      49695 GCCCGGAATTTTCTTTATTTCTATACATTTTACTTCCGTTGTTTTTTTCTTTTT 49748


RESULT 11
ADP13586/c
ID   ADP13586 standard; DNA; 170245 BP.
XX
AC   ADP13586;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Renal cell carcinoma differentially expressed gene #322.
XX
KW   ds; diagnosis; non-blood disease; solid tumor; gene expression;
KW   peripheral blood mononuclear cell; renal cell carcinoma; prostate cancer;
KW   head/neck cancer; differential expression.
XX
OS   Homo sapiens.
XX
PN   WO2004048933-A2.
XX
PD   10-JUN-2004.
XX
PF   21-NOV-2003; 2003WO-US037481.
XX
PR   21-NOV-2002; 2002US-0427982P.
PR   03-APR-2003; 2003US-0459782P.
XX
PA   (AMHP ) WYETH.
PA   (TWIN/) TWINE N C.
PA   (BURC/) BURCZYNSKI M E.
PA   (TREP/) TREPICCHIO W L.
PA   (DORN/) DORNER A.
PA   (STOV/) STOVER J A.
PA   (SLON/) SLONI D K.
XX
PI   Twine NC,  Burczynski ME,  Trepicchio WL,  Dorner A,  Stover JA;
PI   Sloni DK;
XX
DR   WPI; 2004-460799/43.
XX
PT   Diagnosing non-blood disease such as solid tumor, involves comparing
PT   differential expression profile of specific genes in peripheral blood
PT   sample of subject with reference expression profile of specific genes.
XX
PS   Disclosure; SEQ ID NO 322; 350pp; English.
XX
CC   The invention relate to a method of diagnosing (M1) non-blood disease
CC   such as solid tumor by providing peripheral blood sample of human having
CC   non-blood disease, and comparing an expression profile of specific genes
CC   in the peripheral blood sample to reference expression profile of the
CC   genes, where each of the genes is differentially expressed in peripheral
CC   blood mononuclear cells (PBMCs) of patients having the disease as
CC   compared to PBMCs of normal humans. The method is useful for diagnosing
CC   non-blood disease such as solid tumor. The solid tumor is chosen from
CC   renal cell carcinoma (RCC), prostate cancer and head/neck cancer. The
CC   peripheral blood sample comprises enriched PBMCs. The peripheral blood
CC   sample is a whole blood sample (claimed). (M1) is useful for identifying
CC   genes that are differentially expressed in peripheral blood samples
CC   isolated at different stages of progression, development or treatment of
CC   RCC and/or other solid tumors. This sequence corresponds to a gene that
CC   is differentially expressed and detected by the method of the invention.
CC   (Note: this sequence is not given as part of the printed specification
CC   but was obtained from WIPO in electronic format at
CC   ftp.wipo./pub/published_pct_sequences).
XX
SQ   Sequence 170245 BP; 44031 A; 43026 C; 42490 G; 40698 T; 0 U; 0 Other;

  Query Match             10.4%;  Score 208;  DB 12;  Length 170245;
  Best Local Similarity   78.0%;  Pred. No. 1.1e-41;
  Matches  276;  Conservative    0;  Mismatches   75;  Indels    3;  Gaps    2;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| ||||||| || |  ||||||  || ||||||||||||||||||| ||
Db     120850 TTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCA 120791

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| |||  |||| |||||  |||||||||||||||||||||||| ||||||||||
Db     120790 GTGG-TGTAATCTCAGCTCAGCGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGCC 120732

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATT 1393
             ||||| |||| | ||||||||| ||||||| | |||  |  || ||      ||||||||
Db     120731 TCAGCCTCCCAAGTAGCTGGGACTACAGGCACGTGCCACCACACGCGGCTAATTTGTATT 120672

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             |||||||||||||| |||||||||||||| ||||||| | |||||||||||||| |||||
Db     120671 TTTAGTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGG 120612

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||||||||||| ||||||||||  |||||||||||||||||||||| |  |||
Db     120611 TGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCTCCCG 120552

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTT 1567
             ||| |   | |    |  |    |   || |    || |  ||| | |||| ||
Db     120551 GCCCGGAATTTTCTTTATTTCTATACATTTTACTTCCGTTGTTTTTTTCTTTTT 120498


RESULT 12
AAK75677
ID   AAK75677 standard; DNA; 16424 BP.
XX
AC   AAK75677;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:30489.
XX
```

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.

```
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 30489; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 16424 BP; 4399 A; 4136 C; 3644 G; 4245 T; 0 U; 0 Other;

  Query Match           10.3%;  Score 206.8;  DB 4;  Length 16424;
  Best Local Similarity   74.6%;  Pred. No. 8.1e-42;
  Matches  300;  Conservative    0;  Mismatches   97;  Indels    5;  Gaps    3;

Qy       1166 ATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATT 1225
             | ||| |||| |  || | ||| ||    | | ||    |   ||    || | ||
Db       9024 AATTCTGTGTGTATCTTCTAGAGATAATCTATCTAATAAAGAATGGAAGCATATTCTTTT 9083

Qy       1226 TTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGG 1285
             ||||||||||||| |||||||||  || |||||||||||||||| | |||||||
Db       9084 TTTATTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAATGCAGTGGC-ACAA 9142

Qy       1286 TCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGTC--AGCTTCCC 1343
             |||| |||||||||||||||||||||||| |||||| ||||||||||   ||| || |
Db       9143 TCTCAGCTCACTGCAACCTCTGCCTCCCAGTTTCAAGAGATTCTCCTGTCTTAGCCTCTC 9202

Qy       1344 GAATAGCTGGGATTACAGGCG--CATGCACCACCATGCCTAATTTTTGTATTTTTAGTAG 1401
             || ||||||| |||||||||  ||    |   | || ||||||||||||||||||| ||
Db       9203 GAGTAGCTGGAATTACAGGCGTCCACCACCACGCCTGGCTAATTTTTGTATTTTTAGCAG 9262

Qy       1402 AGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCAC 1461
               |||| ||||| |||||||| ||||||| | |||||||||||||| ||||||||||||
Db       9263 TTACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAC 9322

Qy       1462 CCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
             |  |||  |||||||||| |||||||||||||| |||||||| |||||||| || |||
Db       9323 CTGCCTTGGCCTCCCAAAGTACTGGGATTACAGGCGTGAGCCACTGTGCCCAGTCTTTTT 9382

Qy       1522 TCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
             | |    | | ||| |||    |   | |||  || | |
Db       9383 TTTTTTTTTTTTTTTTTTTTTGAGATGGTACATCCCTCTGT 9424


RESULT 13
AAK68448/c
ID   AAK68448 standard; DNA; 16424 BP.
XX
AC   AAK68448;
XX
DT   06-NOV-2001  (first entry)

XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23260.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
```

PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 23260; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 16424 BP; 4245 A; 3644 C; 4136 G; 4399 T; 0 U; 0 Other;

Query Match 10.3%; Score 206.8; DB 4; Length 16424;
Best Local Similarity 74.6%; Pred. No. 8.1e-42;
Matches 300; Conservative 0; Mismatches 97; Indels 5; Gaps 3;

Qy 1166 ATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATT 1225
Db 7401 AATTCTGTGTGTATCTTCTAGAGATAATCTATCTAATAAAGAATGGAAGCATATTCTTTT 7342

Qy 1226 TTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGG 1285
Db 7341 TTTATTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAATGCAGTGGC-ACAA 7283

Qy 1286 TCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGTC--AGCTTCCC 1343
Db 7282 TCTCAGCTCACTGCAACCTCTGCCTCCCAGTTTCAAGAGATTCTCCTGTCTTAGCCTCTC 7223

Qy 1344 GAATAGCTGGGATTACAGGCG--CATGCACCACCATGCCTAATTTTTGTATTTTTAGTAG 1401
Db 7222 GAGTAGCTGGAATTACAGGCGTCCACCACCACGCCTGGCTAATTTTTGTATTTTTAGCAG 7163

Qy 1402 AGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCAC 1461
Db 7162 TTACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAC 7103

Qy 1462 CCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
Db 7102 CTGCCTTGGCCTCCCAAAGTACTGGGATTACAGGCGTGAGCCACTGTGCCCAGTCTTTTT 7043

Qy 1522 TCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
Db 7042 TTTTTTTTTTTTTTTTTTTTTGAGATGGTACATCCCTCTGT 7001

RESULT 14
ADP75184/c
ID ADP75184 standard; DNA; 35192 BP.
XX

AC ADP75184;
XX
DT 12-AUG-2004 (first entry)
XX
DE Human ADAM19 gene fragment #4.
XX
KW Human; chromosome 5; ds; gene; ADAM19; Endophilin 1; Endophilin 2; NRG2;
KW ADAMTS2; a disintegrin and metalloprotease; neuroregulin 2; SNP;
KW single nucleotide polymorphism;
KW a disintegrin and metalloprotease with thrombospondin type1 motif 2;
KW asthma; atopy; obesity; inflammatory bowel disease; respiratory disorder.
XX
OS Homo sapiens.
XX
PN WO2003031594-A2.
XX
PD 17-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032700.
XX
PR 11-OCT-2001; 2001US-0328424P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Keith T, Little RD, Van Eerdewegh P, Dupuis J, Del Mastro RG;
PI Allen K;
XX
DR WPI; 2003-381712/36.
XX
PT New isolated nucleic acid or alternate splice variant, useful for
PT diagnosing and treating a disintegrin and metalloprotease (ADAM) or
PT interactor gene-associated disorder, e.g. asthma, atopy, obesity or
PT inflammatory bowel disease.
XX
PS Claim 2; SEQ ID NO 5; 338pp; English.
XX
CC The invention relates to an isolated nucleic acid or alternate splice
CC variant comprising a nucleotide sequence containing at least one of the
CC single nucleotide polymorphisms given in the specification, a nucleotide
CC sequence having at least 15 contiguous nucleotides of them, or
CC complements of them. The genes are ADAM19 (a disintegrin and
CC metalloprotease 19, also known as gene 845), NRG2 (neuroregulin 2, also
CC known as gene 847), endophilin 1 (also known as gene 874), endophilin 2
CC (also known as gene 803) and ADAMTS2 (a disintegrin and metalloprotease
CC with thrombospondin type1 motif 2, also known as gene 962). Also included
CC are a vector comprising the isolated nucleic acid (or alternate splice
CC variant), a host cell containing the vector, an isolated polypeptide
CC encoded by the novel nucleic acid (or alternate splice variant), an
CC antibody or antibody fragment that binds to the polypeptide,
CC pharmaceutical compositions (comprising the nucleic acid or alternate
CC splice variant, vector, polypeptide or antibody , and a carrier,
CC excipient or diluent), a kit for detecting a disintegrin and
CC metalloprotease (ADAM) gene nucleotide sequence (comprising the isolated
CC nucleic acid or alternate splice variant, antibody or antibody fragment,
CC and at least one component to detect the hybridisation of the variant or
CC the binding of the antibody to an ADAM gene amino acid sequence), a kit
CC for detecting an interactor gene amino acid sequence (comprising the
CC antibody or antibody fragment, and at least one component to detect the
CC binding of the antibody to the interactor gene amino acid sequence),
CC diagnosing an ADAM or interactor gene-associated disorder or a
CC respiratory disorder in a human subject, determining an ADAM or
CC interactor gene pharmacogenetic profile in a human subject, identifying
CC an orthologue of a human ADAM or interactor gene, treating an ADAM or
CC interactor gene-associated disorder (or a respiratory disorder) by
CC administering the pharmaceutical composition, a transgenic mouse (whose
CC genome comprises an introduced null mutation in an endogenous gene that
CC is orthologous to a human ADAM gene), making a homozygous transgenic
CC knockout mouse, forming a crystal of the isolated polypeptide, a cell
CC line comprising the isolated nucleic acid or alternate splice variant, a
CC biochip comprising the isolated nucleic acid or alternate splice variant,
CC an isolated nucleic acid probe or primer comprising at least 8 contiguous
CC nucleotides of the nucleic acid, an isolated antisense nucleic acid,
CC identifying an ADAM or interactor gene ligand and an isolated nucleic
CC acid variant of Gene 803, 845, 847, 874 or 962. The nucleic acid or
CC alternate splice variants, methods, kits and antibody/antibody fragment
CC are useful for diagnosing and treating an ADAM or interactor gene-
CC associated disorder, e.g. asthma, atopy, obesity or inflammatory bowel
CC disease. The present sequence is a gene (or gene fragment) for one of the
CC above mentioned genes. NOTE: This sequence may contain one of the SNPs
CC detailed in table 5 of the specification but their positions cannot be
CC determined using the information given in the patent.
XX
SQ Sequence 35192 BP; 9087 A; 7990 C; 8613 G; 9502 T; 0 U; 0 Other;

Query Match 10.3%; Score 206.8; DB 11; Length 35192;
Best Local Similarity 74.6%; Pred. No. 1.1e-41;
Matches 300; Conservative 0; Mismatches 97; Indels 5; Gaps 3;

Qy 1166 ATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATT 1225
Db 8751 AATTCTGTGTGTATCTTCTAGAGATAATCTATCTAATAAAGAATGGAAGCATATTCTTTT 8692

Qy 1226 TTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGG 1285
Db 8691 TTTATTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAATGCAGTGGC-ACAA 8633

Qy 1286 TCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGTC--AGCTTCCC 1343
Db 8632 TCTCAGCTCACTGCAACCTCTGCCTCCCAGTTTCAAGAGATTCTCCTGTCTTAGCCTCTC 8573

Qy 1344 GAATAGCTGGGATTACAGGCG--CATGCACCACCATGCCTAATTTTTGTATTTTTAGTAG 1401
Db 8572 GAGTAGCTGGAATTACAGGCGTCCACCACCACGCCTGGCTAATTTTTGTATTTTTAGCAG 8513

Qy 1402 AGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCAC 1461
Db 8512 TTACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAC 8453

Qy 1462 CCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
Db 8452 CTGCCTTGGCCTCCCAAAGTACTGGGATTACAGGCGTGAGCCACTGTGCCCAGTCTTTTT 8393

Qy 1522 TCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
Db 8392 TTTTTTTTTTTTTTTTTTTTTTGAGATGGTACATCCCTCTGT 8351

RESULT 15
ABD32861/c
ID ABD32861 standard; DNA; 118466 BP.
XX
AC ABD32861;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD17-042.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW leukaemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.
PR 15-SEP-2003; 2003US-00663431.
PR 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.

```
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 536; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 118466 BP; 30920 A; 25969 C; 28235 G; 33263 T; 0 U; 79 Other;

  Query Match             10.3%;  Score 206.8;  DB 13;  Length 118466;
  Best Local Similarity   74.6%;  Pred. No. 1.8e-41;
  Matches  300;  Conservative    0;  Mismatches   97;  Indels    5;  Gaps    3;

Qy       1166 ATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATT 1225
             | ||| ||||  |  || |  ||| ||    |  | ||    |   ||   || | ||
Db      90189 AATTCTGTGTGTATCTTCTAGAGATAATCTATCTAATAAAGAATGGAAGCATATTCTTTT 90130

Qy       1226 TTTATTTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGG 1285
             |||||||||||||| |||||||||  || |||||||||||||||| | |||||||
Db      90129 TTTATTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAATGCAGTGGC-ACAA 90071

Qy       1286 TCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGTC--AGCTTCCC 1343
             ||||  |||||||||||||||||||||||||| ||||||  ||||||||||||   ||| || |
Db      90070 TCTCAGCTCACTGCAACCTCTGCCTCCCAGTTTCAAGAGATTCTCCTGTCTTAGCCTCTC 90011

Qy       1344 GAATAGCTGGGATTACAGGCG--CATGCACCACCATGCCTAATTTTTGTATTTTTAGTAG 1401
             || ||||||| ||||||||||  ||     |   | || ||||||||||||||||||| ||
Db      90010 GAGTAGCTGGAATTACAGGCGTCCACCACCACGCCTGGCTAATTTTTGTATTTTTAGCAG 89951

Qy       1402 AGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCAC 1461
                |||| ||||| |||||||| ||||||| | ||||||||||||| ||||||||||||
Db      89950 TTACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAC 89891

Qy       1462 CCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
             |  |||  |||||||||| |||||||||||||| |||||||| |||||||| || |||
Db      89890 CTGCCTTGGCCTCCCAAAGTACTGGGATTACAGGCGTGAGCCACTGTGCCCAGTCTTTTT 89831

Qy       1522 TCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
              | |     |  | | ||| |||     |   | |||  || | |
Db      89830 TTTTTTTTTTTTTTTTTTTTTGAGATGGTACATCCCTCTGT 89789


RESULT 16
ACN43946
ID   ACN43946 standard; DNA; 174448 BP.
XX
AC   ACN43946;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG21793.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 148; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 174448 BP; 43106 A; 40529 C; 40481 G; 47948 T; 0 U; 2384 Other;

  Query Match             10.3%;  Score 206.8;  DB 11;  Length 174448;
  Best Local Similarity   79.4%;  Pred. No. 2.2e-41;
  Matches  270;  Conservative    0;  Mismatches   67;  Indels    3;  Gaps    2;

Qy       1180 ATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGA 1239
             | |||  | | |    ||| |   | || |       | ||||  ||||| ||||||||||
Db     115868 AGTGTGTATACATTAGTTTAAAATTATTATTATTATTTTCTTTTTTTTTTTTTTTTGA 115927

Qy       1240 AACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGC 1299
              || ||||||||||  |||||||||||||||| |||||||| || ||| ||||||||||
Db     115928 CACGGAGTCTCACTCCGTCACCCAGGCTGGAGTGCAGTGGCT-TGATCTTGGCTCACTGC 115986

Qy       1300 AACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATT 1357
              |||||| |||| |  |||||||||||||||  ||  | ||| ||| ||||||||||||||
Db     115987 AACCTCCGCCTGCTGGGTTCAAGCGATTCTTGTGCCTTAGCCTCCTGAATAGCTGGGATT 116046

Qy       1358 ACAGGCGCATGCACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCA 1417
```

```
                |||||  ||   |||  ||   | |||||||||||||||||||||||||| | ||||| |||
Db     116047 ACAGGTGCTGCCACTACACCGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCA 116106

Qy       1418 TGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCA 1477
                ||||| ||||| | | |||||| ||||||| |||  ||||||||||  ||||||||||||
Db     116107 TGTTGGCCAGGGTGGTCTTGAATTCCTGACCTCAAATGATCCACCCGTCTCAGCCTCCCA 116166

Qy       1478 AAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCT 1517
                |||  |||||||||||||| ||||||||| |||||||||
Db     116167 AAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCAGCCT 116206


RESULT 17
ADQ97489
ID   ADQ97489 standard; DNA; 48642 BP.
XX
AC   ADQ97489;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer associated sequence HD09-003, SEQ ID 466.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 466; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 48642 BP; 11523 A; 11854 C; 12110 G; 13155 T; 0 U; 0 Other;

  Query Match             10.3%;  Score 205.6;  DB 12;  Length 48642;
  Best Local Similarity   78.8%;  Pred. No. 2.6e-41;
  Matches  271;  Conservative    0;  Mismatches   69;  Indels    4;  Gaps    2;

Qy       1202 TATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACC 1261
              || |   | |||     |||| |||||  ||||| ||| |||||||||||||||||||
Db       8325 TAAGGGATATCATCAGATTTTTTTTTTCCTTTTTCTGAGACAGAGTCTCACTTTGTCACC 8384

Qy       1262 CAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAA 1321
              |||||||||||||||||   | |||||||||||||||||||| ||||| |||||||
Db       8385 CAGGCTGGAGTACAGTGGCGCAGATCTCGGCTCACTGCAACCTCTGGCTCCCGGGTTCAA 8444

Qy       1322 GCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCAT 1377
              | ||||||||||  ||||| ||| || |||||||||||||||||| |||   |   |
Db       8445 GTGATTCTCCTGCCTCAGCCTCCGGAGTAGCTGGGATTACAGGCGCCTGCCACCATGCCC 8504

Qy       1378 GCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTG 1437
              | ||| |||||||||||||||||||| | ||||| |||||||  |||||||   |

Db       8505 GGCTACTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTAGCCAGGCTGTTCCCA 8564

Qy       1438 AACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCA 1497
              |||||||||| |||||  ||||  || ||||||||||||||||  ||||||||||||||
Db       8565 AACTCCTGACCTCAGGCAATCCGGCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCG 8624

Qy       1498 TGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
              |||||||||| |||| ||| |  ||  |   |  | | ||| ||
Db       8625 TGAGCCACCGAGCCCGGCCAGACTTGACTTTTTTTTTTTTTTTT 8668


RESULT 18
AAZ93815/c
ID   AAZ93815 standard; DNA; 144460 BP.
XX
AC   AAZ93815;
XX
DT   16-AUG-2000  (first entry)
XX
DE   Olfactory receptor operon.
XX
KW   Olfactory receptor protein; olfactory; smell; receptor; binding;
KW   detection; screening; genotyping; biallelic marker; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2406. .2600
FT                   /*tag=  a
FT                   /label= ORF1
FT   allele          7521
FT                   /*tag=  m
FT                   /note= "Polymorphic base G or T"
FT   allele          8192
FT                   /*tag=  n
FT                   /note= "Polymorphic base G or T"
FT   CDS             9711. .10658
FT                   /*tag=  b
FT                   /label= ORF2
FT   allele          14483
FT                   /*tag=  o
FT                   /note= "Polymorphic base A or T"
FT   allele          19625
FT                   /*tag=  p
FT                   /note= "Polymorphic base C or T"
FT   allele          20583
FT                   /*tag=  q
FT                   /note= "Polymorphic base A or G"
FT   CDS             24851. .25369
FT                   /*tag=  c
FT                   /label= ORF3
FT   CDS             45714. .46661
FT                   /*tag=  d
FT                   /label= ORF4
FT   allele          76947
FT                   /*tag=  r
FT                   /note= "Polymorphic base A or C"
FT   CDS             80198. .81115
FT                   /*tag=  e
FT                   /label= ORF 5
FT   allele          91088
FT                   /*tag=  s
FT                   /note= "Polymorphic base G or C"
FT   allele          91138
FT                   /*tag=  t
FT                   /note= "Polymorphic base A or G"
FT   allele          91187
FT                   /*tag=  u
FT                   /note= "Polymorphic base A or G"
FT   CDS             96291. .96902
FT                   /*tag=  f
FT                   /label= ORF6
FT   CDS             110758. .111564
```

```
FT                   /*tag=  g
FT                   /label= ORF7
FT   misc_feature    complement(113644. .114063)
FT                   /*tag=  k
FT                   /note= "Ubiquitin 1 pseudogene complement"
FT   CDS             122525. .122887
FT                   /*tag=  h
FT                   /label= ORF8
FT   misc_feature    complement(127489. .127854)
FT                   /*tag=  l
FT                   /note= "Ubiquitin 2 pseudogene complement"
FT   CDS             132454. .133389
FT                   /*tag=  i
FT                   /label= ORF9
FT   allele          133998
FT                   /*tag=  v
FT                   /note= "Polymorphic base A or G"
FT   allele          140066
FT                   /*tag=  w
FT                   /note= "Polymorphic base C or T"
FT   allele          141176
FT                   /*tag=  x
FT                   /note= "Polymorphic base A or G"
FT   CDS             143398. .143577
FT                   /*tag=  j
FT                   /label= ORF10
FT   allele          144033
FT                   /*tag=  y
FT                   /note= "Polymorphic base A or G"
XX
PN   WO200021985-A2.
XX
PD   20-APR-2000.
XX
PF   13-OCT-1999;   99WO-IB001729.
XX
PR   14-OCT-1998;   98US-0104299P.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Malekzadeh K;
XX
DR   WPI; 2000-317933/27.
XX
PT   New nucleic acids encoding ten different olfactory receptor proteins and
PT   their biallelic markers, are useful in genetic analysis and in screening
PT   for compounds which bind to the receptor proteins.
XX
PS   Claim 1; Page 103-141; 155pp; English.
XX
CC   Ten new olfactory receptor proteins and their biallelic markers have been
CC   described. The sequences encoding these receptor proteins and which
CC   contain the biallelic markers can be used for genotyping. The olfactory
CC   receptor proteins can be used to screen for substances which bind to
CC   them. See GENESEQ records AAZ93816-25 and AAY83386-95
XX
SQ   Sequence 144460 BP; 46068 A; 27088 C; 26615 G; 44676 T; 0 U; 13 Other;

  Query Match             10.2%;  Score 204.8;  DB 3;  Length 144460;
  Best Local Similarity   63.8%;  Pred. No. 6.5e-41;
  Matches  394;  Conservative    0;  Mismatches  212;  Indels   12;  Gaps    5;

Qy       1165 CATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTAT 1224
             | ||| ||   ||    |||||   |     ||| ||||||| || |       || ||
Db      57682 CTTTTGAGAAGTGTTTATGTTCTTTGCCCACTTTTATTATGTATTTTATTTATTTATTTA 57623

Qy       1225 TTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTG 1284
              |||  || ||| ||| ||||| ||||||| |||| |||||||||||||| |||||| ||
Db      57622 TTTATTTATTTATGAGACAGAATCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGG-TGCA 57564

Qy       1285 GTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCC 1342
              |||||||||||||||||||| |||||   ||||||| |||||||| |  ||||| ||
Db      57563 ATCTCGGCTCACTGCAACCTCTACCTCCTGAGTTCAAGTGATTCTCCAGCCTCAGCCTCT 57504

Qy       1343 CGAATAGCTGGGATTACAGGCGCATGCACCACCATGCC----TAATTTTTGTATTTTTAG 1398
              || |||||||||||||||  ||| ||||||||||||     || ||||||||||||||||
Db      57503 AGAGTAGCTGGGATTACAG--GCACGCACCACCATGCCCAGTTATTTTTTGTATTTTTAG 57446

Qy       1399 TAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATC 1458
              |||| |||| |||| ||||||||| ||||||| | |||||||||||||| |||| | |||
Db      57445 TAGAAACAGGGTTTTGCCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGATAATC 57386

Qy       1459 CACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTG 1518
              ||||  |||| ||||||||||  |||||||||| |||||||||| ||  ||||| |
Db      57385 CACCTGCCTCTGCCTCCCAAAGTGCTGGGATTACTGGCATGAGCCACTGTTGCCAGCATC 57326

Qy       1519 TTTTCTCAGATCC---TGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTCATTT 1575
              ||  | ||  |     ||| ||  || ||     |||       |    | |  |  |  |
Db      57325 TTAGCCCACTTTTAAATGTGTTGTTTGCTTTTCTCTTGTAAATTTAAGTTTCTTAAACAT 57266

Qy       1576 TGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGGAAAATTCCCTGG 1635
                | || || ||||  || || |   ||       |||||| ||     | | |   ||
Db      57265 GCTGAATATTAGACCTTTGTCAGATACATAGTTTGGAAATATTTTCTTCTATTCCATAGG 57206

Qy       1636 TTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTTATGTTTCTTTCAAAT 1695
              |   || ||| ||    || | | |||| |    |  || | | |      ||    |
Db      57205 CTGTCTTTTTACTCTGTTAACAGTTTGTTTTGCTGTGAAGAAGCTCTTAAGTTTAATTAG 57146

Qy       1696 TATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTATTGAATAGTCTTCACT 1755
                | || || || |  | ||| |||  |  | | |  || |||  |  |  |||
Db      57145 ATCCTACTTGTCAATTTTTGCTTTTTTCTTTTTGCAATTGCTTTTTGTGTCTTTGTCATG 57086

Qy       1756 TCTTGTCATCCAGTTTCT 1773
                |    ||  |||||
Db      57085 AAATCATTGCCCATTTCT 57068


RESULT 19
ACN44262
ID   ACN44262 standard; DNA; 168821 BP.
XX
AC   ACN44262;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG18035.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 622; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
```

```
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 168821 BP; 39588 A; 43389 C; 45655 G; 40189 T; 0 U; 0 Other;

  Query Match             10.2%;  Score 204.8;  DB 11;  Length 168821;
  Best Local Similarity   77.7%;  Pred. No. 6.9e-41;
  Matches  286;  Conservative    0;  Mismatches   77;  Indels    5;  Gaps    3;

Qy       1191 AAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTC 1250
             ||   | |  |||| | ||     || | ||||||| | | |||||||| || | ||||
Db     166269 AATCCTATATTTATTTATTTATTTATTTATTTATTTATTTTTTTGAGACGGTGTCTT 166328

Qy       1251 ACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCT 1310
               || | ||||||||||||||||||||||||     |||| ||||||||||||||  |||
Db     166329 GCTCTATCACCCAGGCTGGAGTACAGTGGC-ACAATCTCAGCTCACTGCAACCTCCACCT 166387

Qy       1311 CCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG 1368
             |||||||||||||||||||||||  ||||| |||||| |||||||||||||||||| |  |
Db     166388 CCCAGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCACCCG 166447

Qy       1369 C--ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCA 1426
             |  |   |  | |||||||||||||||||||||||||||||| ||||| |||||||| |||
Db     166448 CCACCACGCCCGGCTAATTTTTGTATTTTTAGTAGAGACAGTGTTTCACCATGTTGGCCA 166507

Qy       1427 GGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGG 1486
             ||||  | |||||||||||||| ||| | ||||| ||||||||||||||||||||  ||||
Db     166508 GGCTGGTCTTGAACTCCTGACCTCAAGGGATCCGCCCACCTCAGCCTCCCAAAGTGCTGG 166567

Qy       1487 GATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGA 1546
             |||||||||||||||||||| ||||| |||  ||     ||| | | ||| || | |
Db     166568 GATTACAGGCATGAGCCACCATGCCCGGCCCATTACGCAAGAACTTTTATGTGCTCTGTT 166627

Qy       1547 AGCCTTCA 1554
             |  | |||
Db     166628 AATCCTCA 166635


RESULT 20
ABA18360
ID   ABA18360 standard; DNA; 11343 BP.
XX
AC   ABA18360;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 10691.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001334.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
```

```
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   20-OCT-2000; 2000US-0242221P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250391P.
PR   01-DEC-2000; 2000US-0251160P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system cancers
PT   and metastases.
XX
PS   Disclosure; SEQ ID NO 10691; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 11343 BP; 3085 A; 2464 C; 2534 G; 3260 T; 0 U; 0 Other;

  Query Match           10.2%;  Score 203.4;  DB 5;  Length 11343;
  Best Local Similarity   70.0%;  Pred. No. 5.2e-41;
  Matches  317;  Conservative    0;  Mismatches  131;  Indels    5;  Gaps    3;

Qy       1125 TATAAGGAGATTTTTTAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGT 1184
             | ||| ||  |  | | |     |||  | | |||  ||||||   | |    | ||
Db       2027 TTTAAAGACCCTGGTAAACATCTGTTGAACATAAAGATACCATTTGCCTTTCAGGACTGG 2086

Qy       1185 TCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAG 1244
              | ||   |   |  | || | | |  | |  |    || ||||| |||| ||||| ||||
Db       2087 TTAGTATACAGTAGGTTTGTTTCTAATTATGTTGGGTTTTTTTTGTTTTGTTTGAGACAG 2146

Qy       1245 AGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCT 1304
              ||||||||| |||| |||||||||||| ||||||| |   ||| || ||||||||||||
Db       2147 AGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGC-GCCATCTTGGATCACTGCAACCT 2205

Qy       1305 CTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGG 1362
              |||||||||||||||||||| |||||||  ||||| |||| | ||||||||||||||||
Db       2206 CTGCCTCCCAGGTTCAAGCGAGTCTCCTGCCTCAGCCTCCCTAGTAGCTGGGATTACAGG 2265

Qy       1363 CGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGT 1420
               |||| |  |  |   |  |||||||||||||||||||||||   | ||||| ||||||
Db       2266 TGCATACCACCACGCCCAGCTAATTTTTGTATTTTTAGTAGAGGTGGGGTTTCACCATGT 2325

Qy       1421 TGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAG 1480
              || ||||||| | || |||||||||| ||| |||||||||| |||||||||| ||||
Db       2326 TGGCCAGGCTGGTCTCGAACTCCTGACCTCAAGTGATCCACCCGCCTCAGCCTCCTAAAG 2385

Qy       1481 CACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGT 1540
               |||||||||||||  |||||||| |||||||||| |      |   |  |      |
Db       2386 GGCTGGGATTACAGGTGTGAGCCACTGTGCCCAGCCAGGGCAAACTTTTTTTTAGCAGGA 2445

Qy       1541 TTCTGAAGCCTTCATTTCTATCTTCTTATTCAT 1573
               |   |   |  ||  || ||| | || ||| |
Db       2446 ATTATACCTCACCAGATCCATCATTTTTTTCTT 2478


RESULT 21
ADC21029
ID   ADC21029 standard; DNA; 11343 BP.
XX
AC   ADC21029;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human secreted protein-related DNA sequence #447.
XX
KW   gene therapy; human; secreted protein; haemopoietic disorder;
KW   haematological disorder; anaemia; haemophilia; inflammatory disorder;
KW   inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
KW   leukaemia; wound healing; epithelial cell proliferation disorder;
KW   immune disorder; autoimmune disorder; asthmatic disorder;
KW   cardiovascular disorder; atherosclerosis; myocarditis;
```

KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200292787-A2.
XX
PD 21-NOV-2002.
XX
PF 26-MAR-2002; 2002WO-US009257.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-129287/12.
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating hematopoietic or hematologic disorders, e.g.
PT anemia or hemophilia.
XX
PS Disclosure; SEQ ID NO 983; 1512pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC wound healing and disorders of epithelial cell proliferation; immune
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence was used in the
CC exemplification of the invention.
XX
SQ Sequence 11343 BP; 3085 A; 2464 C; 2534 G; 3260 T; 0 U; 0 Other;

Query Match 10.2%; Score 203.4; DB 10; Length 11343;
Best Local Similarity 70.0%; Pred. No. 5.2e-41;
Matches 317; Conservative 0; Mismatches 131; Indels 5; Gaps 3;

Qy 1125 TATAAGGAGATTTTTTAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGT 1184
Db 2027 TTTAAAGACCCTGGTAAACATCTGTTGAACATAAAGATACCATTTGCCTTTCAGGACTGG 2086

Qy 1185 TCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAG 1244
Db 2087 TTAGTATACAGTAGGTTTGTTTCTAATTATGTTGGGTTTTTTTTGTTTTGTTTGAGACAG 2146

Qy 1245 AGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCT 1304
Db 2147 AGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGC-GCCATCTTGGATCACTGCAACCT 2205

Qy 1305 CTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGG 1362
Db 2206 CTGCCTCCCAGGTTCAAGCGAGTCTCCTGCCTCAGCCTCCCTAGTAGCTGGGATTACAGG 2265

Qy 1363 CGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGT 1420
Db 2266 TGCATACCACCACGCCCAGCTAATTTTTGTATTTTTAGTAGAGGTGGGGTTTCACCATGT 2325

Qy 1421 TGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAG 1480
Db 2326 TGGCCAGGCTGGTCTCGAACTCCTGACCTCAAGTGATCCACCCGCCTCAGCCTCCTAAAG 2385

Qy 1481 CACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGT 1540
Db 2386 GGCTGGGATTACAGGTGTGAGCCACTGTGCCCAGCCAGGGCAAACTTTTTTTTAGCAGGA 2445

Qy 1541 TTCTGAAGCCTTCATTTCTATCTTCTTATTCAT 1573
Db 2446 ATTATACCTCACCAGATCCATCATTTTTTTCTT 2478

RESULT 22
ABT17025
ID ABT17025 standard; DNA; 11343 BP.
XX
AC ABT17025;
XX
DT 03-APR-2003 (first entry)
XX
DE Human secreted protein-related DNA sequence - SEQ ID No 379.
XX
KW Human; gene; ds; protein therapy; immediate hypersensitivity disease;
KW allergic disorder; asthmatic disorder; gene therapy; secreted protein;
KW hay fever; allergic conjunctivitis; allergic rhinitis;
KW binding partner identification; chromosome identification;
KW radiation hybrid mapping; long-range restriction mapping.
XX
OS Homo sapiens.
XX
PN WO200277188-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-US009239.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-175010/17.
XX
PT Use of human secreted proteins and nucleic acids for preparing a
PT diagnostic or pharmaceutical composition for diagnosing or treating
PT allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic
PT conjunctivitis or rhinitis.
XX
PS Disclosure; Page 820-823; 823pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for the diagnosis and treatment of allergic disorders, asthmatic
CC disorders and immediate hypersensitivity diseases (e.g. hay fever,
CC allergic conjunctivitis and allergic rhinitis). The proteins of the
CC invention are also useful for identifying a binding partner. The nucleic
CC acids of the invention are also useful for chromosome identification,
CC radiation hybrid mapping or long-range restriction mapping. The present
CC DNA sequence represents a human secreted protein-related DNA sequence
XX
SQ Sequence 11343 BP; 3085 A; 2464 C; 2534 G; 3260 T; 0 U; 0 Other;

Query Match 10.2%; Score 203.4; DB 10; Length 11343;
Best Local Similarity 70.0%; Pred. No. 5.2e-41;
Matches 317; Conservative 0; Mismatches 131; Indels 5; Gaps 3;

Qy 1125 TATAAGGAGATTTTTTAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGT 1184
Db 2027 TTTAAAGACCCTGGTAAACATCTGTTGAACATAAAGATACCATTTGCCTTTCAGGACTGG 2086

Qy 1185 TCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAG 1244
Db 2087 TTAGTATACAGTAGGTTTGTTTCTAATTATGTTGGGTTTTTTTTGTTTTGTTTGAGACAG 2146

```
Qy       1245 AGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCT 1304
             ||||||||| |||| ||||||||||||| ||||||| |   ||| || ||||||||||||
Db       2147 AGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGC-GCCATCTTGGATCACTGCAACCT 2205

Qy       1305 CTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGG 1362
             ||||||||||||||||||||| |||||||  ||||| |||| | ||||||||||||||||
Db       2206 CTGCCTCCCAGGTTCAAGCGAGTCTCCTGCCTCAGCCTCCCTAGTAGCTGGGATTACAGG 2265

Qy       1363 CGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGT 1420
              |||| |  |  |   |  |||||||||||||||||||||||   | ||||| ||||||
Db       2266 TGCATACCACCACGCCCAGCTAATTTTTGTATTTTTAGTAGAGGTGGGGTTTCACCATGT 2325

Qy       1421 TGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAG 1480
             || ||||||| | || |||||||||||| ||| |||||||||||| |||||||||||| ||||
Db       2326 TGGCCAGGCTGGTCTCGAACTCCTGACCTCAAGTGATCCACCCGCCTCAGCCTCCTAAAG 2385

Qy       1481 CACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGT 1540
               |||||||||||||  |||||||| |||||||||| |       |   | |       |
Db       2386 GGCTGGGATTACAGGTGTGAGCCACTGTGCCCAGCCAGGGCAAACTTTTTTTTAGCAGGA 2445

Qy       1541 TTCTGAAGCCTTCATTTCTATCTTCTTATTCAT 1573
               |   |   |  ||  || ||| | || ||| |
Db       2446 ATTATACCTCACCAGATCCATCATTTTTTTCTT 2478
```

RESULT 23
ABZ68170
ID ABZ68170 standard; DNA; 11343 BP.
XX
AC ABZ68170;
XX
DT 26-MAR-2003 (first entry)
XX
DE Human secreted protein encoding genomic DNA SEQ ID NO 1693.
XX
KW Human; secreted protein; nootropic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW vulnerary; antibacterial; antiparkinsonian; antisickling; antianaemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277186-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-US009188.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040583/03.
XX
PT New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
PS Disclosure; Page 2337-2339; 2423pp; English.
XX
CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 11343 BP; 3085 A; 2464 C; 2534 G; 3260 T; 0 U; 0 Other;

Query Match 10.2%; Score 203.4; DB 10; Length 11343;
Best Local Similarity 70.0%; Pred. No. 5.2e-41;
Matches 317; Conservative 0; Mismatches 131; Indels 5; Gaps 3;

```
Qy       1125 TATAAGGAGATTTTTTAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGT 1184
             | ||| ||   |  | | |      |||  | | |||  ||||||   | |    | ||
Db       2027 TTTAAAGACCCTGGTAAACATCTGTTGAACATAAAGATACCATTTGCCTTTCAGGACTGG 2086

Qy       1185 TCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAG 1244
             | ||   |   | | || | | |  | |  |    || ||||| |||| ||||| ||||
Db       2087 TTAGTATACAGTAGGTTTGTTTCTAATTATGTTGGGTTTTTTTTGTTTTGTTTGAGACAG 2146

Qy       1245 AGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCT 1304
             ||||||||| |||| ||||||||||||| ||||||| |   ||| || ||||||||||||
Db       2147 AGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGC-GCCATCTTGGATCACTGCAACCT 2205

Qy       1305 CTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGG 1362
             ||||||||||||||||||||| |||||||  ||||| |||| | ||||||||||||||||
Db       2206 CTGCCTCCCAGGTTCAAGCGAGTCTCCTGCCTCAGCCTCCCTAGTAGCTGGGATTACAGG 2265

Qy       1363 CGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGT 1420
              |||| |  |  |   |  |||||||||||||||||||||||   | ||||| ||||||
Db       2266 TGCATACCACCACGCCCAGCTAATTTTTGTATTTTTAGTAGAGGTGGGGTTTCACCATGT 2325

Qy       1421 TGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAG 1480
             || ||||||| | || |||||||||||| ||| |||||||||||| |||||||||||| ||||
Db       2326 TGGCCAGGCTGGTCTCGAACTCCTGACCTCAAGTGATCCACCCGCCTCAGCCTCCTAAAG 2385

Qy       1481 CACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGT 1540
               |||||||||||||  |||||||| |||||||||| |       |   | |       |
Db       2386 GGCTGGGATTACAGGTGTGAGCCACTGTGCCCAGCCAGGGCAAACTTTTTTTTAGCAGGA 2445

Qy       1541 TTCTGAAGCCTTCATTTCTATCTTCTTATTCAT 1573
               |   |   |  ||  || ||| | || ||| |
Db       2446 ATTATACCTCACCAGATCCATCATTTTTTTCTT 2478
```

RESULT 24
ACN44586/c
ID ACN44586 standard; DNA; 118931 BP.
XX
AC ACN44586;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG30014.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX

```
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 1108; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 118931 BP; 33837 A; 24131 C; 23530 G; 35376 T; 0 U; 2057 Other;

  Query Match             10.1%;  Score 202.8;  DB 11;  Length 118931;
  Best Local Similarity   83.6%;  Pred. No. 1.9e-40;
  Matches  266;  Conservative    0;  Mismatches   47;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             ||||| | ||||| |||||||||| |||||||||||| |||  |||||||| |||||||
Db      76898 TCTTTCTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTTGCCCAGGCTAGAGTACA 76839

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| || | ||||||||||||||||||||||| |||||||| ||||||||||||||
Db      76838 GTGG-TGCGATCTCGGCTCACTGCAACCTCTGCTTCCCAGGTCCAAGCGATTCTCCTGCC 76780

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTA 1391
             ||||| |||||| ||||||||| || |||| | |||  |  ||  || ||||||| ||||
Db      76779 TCAGCCTCCCGAGTAGCTGGGACTATAGGCACGTGCCACCACACCTGGCTAATTTCTGTA 76720

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             |||||||||||||||| |||| |||||||| ||||||| | || ||||||||||||||||
Db      76719 TTTTTAGTAGAGACAGGGTTTTACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACTTCA 76660

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
             ||||||||||   ||  ||||||||||   |||||||||||||||||||||||||||||
Db      76659 GGTGATCCACCTGTCTTGGCCTCCCAAAGTGTTGGGATTACAGGCATGAGCCACCGTGCC 76600

Qy       1512 CAGCCTGTTTTCTCAGAT 1529
              | |||  |   |||  ||
Db      76599 CGGCCCCTGCACTCCTAT 76582


RESULT 25
ADH19244
ID   ADH19244 standard; DNA; 124164 BP.
XX
AC   ADH19244;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Human 5-hydroxytryptamine receptor type 3B SNP variant DNA.
XX
KW   HTR3B; 5-hydroxytryptamine receptor type 3B; 5-HT; antiemetic;
KW   tranquiliser; neuroleptic; antialcoholic; antimigraine; analgesic;
KW   gastrointestinal; setrone; central nervous system; drug treatment;
KW   postoperative nausea; vomiting; chronic fatigue syndrome;
KW   obsessive-compulsive disorder; schizophrenia; alcoholism; anxiety;
KW   tourette syndrome; migraine; headache; gastrointestinal motility;
KW   cancer chemotherapy; forensic marker; ds; human; SNP;
KW   single nucleotide polymorphism; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003097873-A2.
XX
PD   27-NOV-2003.
XX
PF   15-MAY-2003; 2003WO-EP005120.
XX
PR   15-MAY-2002; 2002EP-00010209.
XX
PA   (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI   Brockmoeller HJ;
XX
DR   WPI; 2004-022892/02.
XX
PT   New 5-Hydroxytryptamine receptor type 3B polynucleotide, useful for
PT   diagnosing and/or treating a setrone-treatable disease such as disorders
PT   of central and/or peripheral nervous system e.g., schizophrenia.
XX
PS   Claim 1; Page; 150pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding an HTR3B (5-
CC   hydroxytryptamine [5-HT] receptor type 3B) polypeptide or fragment having
CC   an amino acid substitution. The polynucleotide of the invention
CC   demonstrates antiemetic, tranquiliser, neuroleptic, antialcoholic,
CC   antimigraine, analgesic and gastrointestinal activities and may be useful
CC   in preparing a composition for diagnosing or treating a disease,
CC   particularly a setrone-treatable disease. Such a disease or dysregulation
CC   is related to the central and peripheral nervous system or secondary to
CC   drug treatment, such as postoperative nausea and/or vomiting, chronic
CC   fatigue syndrome, obsessive-compulsive disorders, schizophrenia,
CC   alcoholism, anxiety disorders, tourette syndrome, migraine, headache and
CC   gastrointestinal motility disorders, preferably nausea and/or vomiting
CC   secondary to cancer chemotherapy. The polynucleotides and polypeptides
CC   may also be useful as forensic markers. The current sequence is that of
CC   the human HTR3B SNP variant DNA of the invention in which bases 36678-
CC   36680 of the alternative SNP form have been deleted. This sequence is not
CC   shown in the specification per se but was created by the indexer using
CC   information from claim 1.
XX
SQ   Sequence 124164 BP; 32694 A; 28609 C; 27969 G; 34892 T; 0 U; 0 Other;

  Query Match             10.1%;  Score 202.6;  DB 12;  Length 124164;
  Best Local Similarity   84.3%;  Pred. No. 2.2e-40;
  Matches  264;  Conservative    0;  Mismatches   44;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| |||||||||| |  |||||||||||||| |||||||||||||| ||
Db      16571 TTTTTTTTTTTTTTTTTTTTTGAGAGGGAGTCTCACTTTGTGACCCAGGCTGGAGTGCA 16630

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| || | |||| ||||||  |||||||||||| | ||||||||||||||||||||
Db      16631 GTGG-TGCGATCTCAGCTCACCACAACCTCTGCCTCTCTGGTTCAAGCGATTCTCCTGCC 16689

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTA 1391
             ||||| ||| || || |||||||||||| || |||   |   | || ||||||||||||
Db      16690 TCAGCCTCCTGAGTAACTGGGATTACAGGTGCCTGCCACCATACCTGGCTAATTTTTGTA 16749

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             |||||||||||||| | ||||| |||||||| ||||||| || ||||||||| ||| |||
Db      16750 TTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGCGTTGAACTCCCGACCTCA 16809

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
             ||||||||||||||||  ||||||||||  |||||||||||||||||||||| |  ||
Db      16810 GGTGATCCACCCACCTCGTCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCCCC 16869

Qy       1512 CAGCCTGTTTTCT 1524
              |||||||||  ||
Db      16870 CAGCCTGTTACCT 16882
```

```
RESULT 26
ADH19246
ID   ADH19246 standard; DNA; 124165 BP.
XX
AC   ADH19246;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Human 5-hydroxytryptamine receptor type 3B SNP variant DNA 2.
XX
KW   HTR3B; 5-hydroxytryptamine receptor type 3B; 5-HT; antiemetic;
KW   tranquiliser; neuroleptic; antialcoholic; antimigraine; analgesic;
KW   gastrointestinal; setrone; central nervous system; drug treatment;
KW   postoperative nausea; vomiting; chronic fatigue syndrome;
KW   obsessive-compulsive disorder; schizophrenia; alcoholism; anxiety;
KW   tourette syndrome; migraine; headache; gastrointestinal motility;
KW   cancer chemotherapy; forensic marker; ds; human; SNP;
KW   single nucleotide polymorphism; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003097873-A2.
XX
PD   27-NOV-2003.
XX
PF   15-MAY-2003; 2003WO-EP005120.
XX
PR   15-MAY-2002; 2002EP-00010209.
XX
PA   (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI   Brockmoeller HJ;
XX
DR   WPI; 2004-022892/02.
XX
PT   New 5-Hydroxytryptamine receptor type 3B polynucleotide, useful for
PT   diagnosing and/or treating a setrone-treatable disease such as disorders
PT   of central and/or peripheral nervous system e.g., schizophrenia.
XX
PS   Claim 1; Page; 150pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding an HTR3B (5-
CC   hydroxytryptamine [5-HT] receptor type 3B) polypeptide or fragment having
CC   an amino acid substitution. The polynucleotide of the invention
CC   demonstrates antiemetic, tranquiliser, neuroleptic, antialcoholic,
CC   antimigraine, analgesic and gastrointestinal activities and may be useful
CC   in preparing a composition for diagnosing or treating a disease,
CC   particularly a setrone-treatable disease. Such a disease or dysregulation
CC   is related to the central and peripheral nervous system or secondary to
CC   drug treatment, such as postoperative nausea and/or vomiting, chronic
CC   fatigue syndrome, obsessive-compulsive disorders, schizophrenia,
CC   alcoholism, anxiety disorders, tourette syndrome, migraine, headache and
CC   gastrointestinal motility disorders, preferably nausea and/or vomiting
CC   secondary to cancer chemotherapy. The polynucleotides and polypeptides
CC   may also be useful as forensic markers. The current sequence is that of
CC   the human HTR3B SNP variant DNA 2 of the invention in which bases 64500-
CC   64501 of the alternative SNP form have been deleted. This sequence is not
CC   shown in the specification per se but was created by the indexer using
CC   information from claim 1.
XX
SQ   Sequence 124165 BP; 32695 A; 28608 C; 27970 G; 34892 T; 0 U; 0 Other;

  Query Match             10.1%;  Score 202.6;  DB 12;  Length 124165;
  Best Local Similarity   84.3%;  Pred. No. 2.2e-40;
  Matches  264;  Conservative    0;  Mismatches   44;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| |||||||||| |  |||||||||||||| |||||||||||||| ||
Db      16571 TTTTTTTTTTTTTTTTTTTTTTTGAGAGGGAGTCTCACTTTGTGACCCAGGCTGGAGTGCA 16630

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
                |||| || | |||| ||||||  |||||||||||| | ||||||||||||||||||
Db      16631 GTGG-TGCGATCTCAGCTCACCACAACCTCTGCCTCTCTGGTTCAAGCGATTCTCCTGCC 16689

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTA 1391
             ||||| ||| || || |||||||||||| || |||    |   | || ||||||||||||
Db      16690 TCAGCCTCCTGAGTAACTGGGATTACAGGTGCCTGCCACCATACCTGGCTAATTTTTGTA 16749

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             |||||||||||||| | ||||| |||||||| ||||||| || ||||||||| ||| |||
Db      16750 TTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGCGTTGAACTCCCGACCTCA 16809

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
             |||||||||||||||||  ||||||||||  |||||||||||||||||||||||| |  ||
Db      16810 GGTGATCCACCCACCTCGTCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCCCC 16869

Qy       1512 CAGCCTGTTTTCT 1524
              |||||||||  ||
Db      16870 CAGCCTGTTACCT 16882


RESULT 27
ADF83402
ID   ADF83402 standard; DNA; 124167 BP.
XX
AC   ADF83402;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Human 5-hydroxytryptamine receptor type 3 gene.
XX
KW   Human; antiemetic; setrone; 5-hydroxytryptamine receptor type 3;
KW   receptor; single nucleotide polymorphism; SNP; HTR3B; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003100091-A1.
XX
PD   04-DEC-2003.
XX
PF   22-MAY-2003; 2003WO-EP005366.
XX
PR   24-MAY-2002; 2002EP-00011491.
XX
PA   (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI   Brockmoeller HJ;
XX
DR   WPI; 2004-035165/03.
DR   GENBANK; AP001874.
XX
PT   Use of setrones for preparing a pharmaceutical composition for treating
PT   or preventing setrone-treatable diseases in a subject having in its
PT   genome less than three copies of a polynucleotide encoding a functional
PT   CYP2D6 polypeptide.
XX
PS   Disclosure; SEQ ID NO 52; 153pp; English.
XX
CC   The present sequence comprises the human 5-hydroxytryptamine receptor
CC   type 3 HTR3B gene. In a variant of the gene ADF83375, nucleotides 36678-
CC   36680 (AAG) of this sequence are deleted. The invention relates to the
CC   use of setrones (antiemetics) for treating and/or preventing setrone-
CC   treatable diseases in a subject having in its genome fewer than 3 copies
CC   of a polynucleotide encoding a functional CYP2D6 polypeptide, and also
CC   having in its genome a second variant allele comprising a polynucleotide
CC   capable of hybridising to a HTR3B gene having AGG at positions 36678-
CC   36680. The treatment regimen can be modified according to the genotype of
CC   the subject's CYP2D6 and/or HTR3B gene. Non-responders to antiemetic
CC   therapy can be identified on a pharmacogenetic basis, allowing a suitable
CC   therapy to be selected. The setrone-treatable diseases are postoperative
CC   nausea and/or vomiting, or nausea and/or vomiting secondary to cancer
CC   chemotherapy, radiation therapy, migraine, acetaminophen poisoning,
CC   prostacyclin therapy, and opioid treatment, spinal or epidural opioid-
CC   related pruritus, acute levodopa-induced psychosis, bulimia nervosa,
```

```
CC   fibromyalgia, chronic fatigue syndrome, obsessive-compulsive disorders,
CC   schizophrenia, alcoholism, cocaine addiction, opioid withdrawal syndrome,
CC   drug withdrawal phenomena, anxiety disorders, cognitive disturbances,
CC   neuroleptic-induced tardive dyskinesia, Tourette's syndrome, migraine
CC   headache or gastrointestinal motility disorder (all claimed).
XX
SQ   Sequence 124167 BP; 32696 A; 28609 C; 27970 G; 34892 T; 0 U; 0 Other;

  Query Match             10.1%;  Score 202.6;  DB 12;  Length 124167;
  Best Local Similarity   84.3%;  Pred. No. 2.2e-40;
  Matches  264;  Conservative    0;  Mismatches   44;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| |||||||||| |  |||||||||||||| |||||||||||||| ||
Db      16571 TTTTTTTTTTTTTTTTTTTTTTGAGAGGGAGTCTCACTTTGTGACCCAGGCTGGAGTGCA 16630

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| || | |||| ||||||  |||||||||||||| | ||||||||||||||||||
Db      16631 GTGG-TGCGATCTCAGCTCACCACAACCTCTGCCTCTCTGGTTCAAGCGATTCTCCTGCC 16689

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTA 1391
             ||||| ||| || || |||||||||||||| || |||   |   | || ||||||||||||
Db      16690 TCAGCCTCCTGAGTAACTGGGATTACAGGTGCCTGCCACCATACCTGGCTAATTTTTGTA 16749

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             |||||||||||||| | ||||| |||||||| ||||||| || |||||||||| ||| |||
Db      16750 TTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGCGTTGAACTCCCGACCTCA 16809

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
             |||||||||||||||||  ||||||||||  |||||||||||||||||||||||| |  ||
Db      16810 GGTGATCCACCCACCTCGTCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCCCC 16869

Qy       1512 CAGCCTGTTTTCT 1524
             |||||||||  ||
Db      16870 CAGCCTGTTACCT 16882


RESULT 28
ADH19240
ID   ADH19240 standard; DNA; 124167 BP.
XX
AC   ADH19240;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Human 5-hydroxytryptamine receptor type 3B gDNA.
XX
KW   HTR3B; 5-hydroxytryptamine receptor type 3B; 5-HT; antiemetic;
KW   tranquiliser; neuroleptic; antialcoholic; antimigraine; analgesic;
KW   gastrointestinal; setrone; central nervous system; drug treatment;
KW   postoperative nausea; vomiting; chronic fatigue syndrome;
KW   obsessive-compulsive disorder; schizophrenia; alcoholism; anxiety;
KW   tourette syndrome; migraine; headache; gastrointestinal motility;
KW   cancer chemotherapy; forensic marker; ds; human; SNP;
KW   single nucleotide polymorphism; gene.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(36777,A)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63724,T)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63725,G)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64152,C)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64228,A)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64757,T)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64910,A)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(65011,A)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(74691,T)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(74737,A)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   WO2003097873-A2.
XX
PD   27-NOV-2003.
XX
PF   15-MAY-2003; 2003WO-EP005120.
XX
PR   15-MAY-2002; 2002EP-00010209.
XX
PA   (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI   Brockmoeller HJ;
XX
DR   WPI; 2004-022892/02.
DR   P-PSDB; ADH19241, ADH19243.
XX
PT   New 5-Hydroxytryptamine receptor type 3B polynucleotide, useful for
PT   diagnosing and/or treating a setrone-treatable disease such as disorders
PT   of central and/or peripheral nervous system e.g., schizophrenia.
XX
PS   Claim 1; SEQ ID NO 49; 150pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding an HTR3B (5-
CC   hydroxytryptamine [5-HT] receptor type 3B) polypeptide or fragment having
CC   an amino acid substitution. The polynucleotide of the invention
CC   demonstrates antiemetic, tranquiliser, neuroleptic, antialcoholic,
CC   antimigraine, analgesic and gastrointestinal activities and may be useful
CC   in preparing a composition for diagnosing or treating a disease,
CC   particularly a setrone-treatable disease. Such a disease or dysregulation
CC   is related to the central and peripheral nervous system or secondary to
CC   drug treatment, such as postoperative nausea and/or vomiting, chronic
CC   fatigue syndrome, obsessive-compulsive disorders, schizophrenia,
CC   alcoholism, anxiety disorders, tourette syndrome, migraine, headache and
CC   gastrointestinal motility disorders, preferably nausea and/or vomiting
CC   secondary to cancer chemotherapy. The polynucleotides and polypeptides
CC   may also be useful as forensic markers. The current sequence is that of
CC   the human HTR3B gDNA of the invention.
XX
SQ   Sequence 124167 BP; 32696 A; 28609 C; 27970 G; 34892 T; 0 U; 0 Other;

  Query Match             10.1%;  Score 202.6;  DB 12;  Length 124167;
  Best Local Similarity   84.3%;  Pred. No. 2.2e-40;
  Matches  264;  Conservative    0;  Mismatches   44;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| |||||||||| |  |||||||||||||| |||||||||||||| ||
Db      16571 TTTTTTTTTTTTTTTTTTTTTTGAGAGGGAGTCTCACTTTGTGACCCAGGCTGGAGTGCA 16630

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| || | |||| ||||||  |||||||||||||| | ||||||||||||||||||
Db      16631 GTGG-TGCGATCTCAGCTCACCACAACCTCTGCCTCTCTGGTTCAAGCGATTCTCCTGCC 16689

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTA 1391
             ||||| ||| || || |||||||||||||| || |||   |   | || ||||||||||||
Db      16690 TCAGCCTCCTGAGTAACTGGGATTACAGGTGCCTGCCACCATACCTGGCTAATTTTTGTA 16749

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
```

```
              ||||||||||||||| | ||||| |||||||| ||||||| || ||||||||| ||| |||
Db      16750 TTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGCGTTGAACTCCCGACCTCA 16809

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
              ||||||||||||||||  |||||||||  |||||||||||||||||||||| |  ||
Db      16810 GGTGATCCACCCACCTCGTCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCCCC 16869

Qy       1512 CAGCCTGTTTTCT 1524
              |||||||||  ||
Db      16870 CAGCCTGTTACCT 16882


RESULT 29
AAK70510
ID   AAK70510 standard; DNA; 6670 BP.
XX
AC   AAK70510;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25322.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
```

```
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 25322; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 6670 BP; 1577 A; 1614 C; 1756 G; 1720 T; 0 U; 3 Other;

  Query Match            10.1%;  Score 202.4;  DB 4;  Length 6670;
  Best Local Similarity   85.3%;  Pred. No. 7.4e-41;
  Matches  261;  Conservative    0;  Mismatches   41;  Indels    4;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| |||||||||||  |||||| ||||| |||||||||||||||||| ||
Db       2519 TTTTTTTTTTTTTTTTTTTTTTGAGGCAGAGTTTCACTCTGTCACCCAGGCTGGAGTGCA 2578

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
              |||||   | ||| |||||||| |||||||||||||||||||||||||||||||||
Db       2579 GTGGC-ACGATCTTGGCTCACTACAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGCC 2637

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC-CATGCCTAATTTTTGTAT 1392
              ||||| |||||| ||||||| ||||||||| |  ||||||| | || ||||||||| |||
Db       2638 TCAGCCTCCCGAGTAGCTGGAATTACAGGCACCCGCACCACGCCTGGCTAATTTTTATAT 2697

Qy       1393 TTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAG 1452
              |||||||||||| | ||||| |||||||| ||||||||  || |||||||||| | ||

Db       2698 TTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTTTTCTCGAACTCCTGACCTTAG 2757

Qy       1453 GTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCC 1512
              ||||||| |||||||| |||||||||||  |||||||||||||  |||||||||   |||
Db       2758 GTGATCCGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCACACCC 2817

Qy       1513 AGCCTG 1518
              ||||||
Db       2818 GGCCTG 2823


RESULT 30
AEA61181/c
ID   AEA61181 standard; DNA; 64829 BP.
XX
AC   AEA61181;
XX
DT   25-AUG-2005  (first entry)
XX
DE   Human FHL1 gene genomic sequence SEQ ID NO:91.
XX
KW   DNA methylation; biomarker; cancer; gene; ds; FHL1.
XX
OS   Homo sapiens.
XX
PN   US2005130172-A1.
XX
PD   16-JUN-2005.
XX
PF   27-JAN-2004; 2004US-00765790.
XX
PR   16-DEC-2003; 2003US-00737082.
XX
PA   (FARB ) BAYER CORP.
XX
PI   Beard C,  Burgess C,  Gannon A,  Harvey J,  Lechner JF,  Li Z;
XX
DR   WPI; 2005-456991/46.
DR   GENBANK; AF098518, AF220153, U29538, NM_001449.
XX
PT   Identifying nucleic acid sequences as biomarker for disease, by
PT   identifying nucleic acid sequences comprising methylated CpG site and
PT   down-regulated in diseased cells and comparing its expression level with
PT   demethylated nucleic acid.
XX
PS   Claim 11; SEQ ID NO 91; 27pp; English.
XX
CC   The invention relates to a method (M1) for identifying one or more
CC   nucleic acid sequences useful as a biomarker for a disease to be
CC   detected. (M1) involves identifying nucleic acid sequences comprising
CC   methylated CpG site in promoter-first exon region and that are down-
CC   regulated in diseased cells, comparing expression level of nucleic acid
CC   sequences with that of demethylated nucleic acid sequences and
CC   identifying nucleic acid sequences exhibiting increase in expression
CC   after demethylation. Also described: (1) detecting (M2) the presence or
CC   stage of a disease in a subject, which involves determining the degree of
CC   methylation of one or more CpG sites on nucleic acid sequences in a
CC   biological sample obtained from the subject, and determining the presence
CC   of, predisposition to, or stage of the disease in the subject based on
CC   the degree of methylation; (2) monitoring the onset, progression, or
CC   regression of a disease in a subject; (3) determining the efficacy of a
CC   test compound for inhibiting a disease in a subject; and (4) a kit (I)
CC   useful for diagnosis, prognosis, staging, monitoring, and therapeutic
CC   treatment of a disease. (M1) is useful for identifying one or more
CC   nucleic acid sequences useful as a biomarker for a disease to be
CC   detected, where the nucleic acid sequences are useful for detecting, the
CC   presence or stage of a disease such as cancer e.g. colorectal cancer in a
CC   subject. The present sequence represents a specifically claimed human
CC   genomic sequence for use in the method of the invention. Note - The
CC   sequence data for this patent is not represented in the printed
CC   specification but was obtained in electronic format from the USPTO web
CC   site.
XX
```

```
SQ   Sequence 64829 BP; 17935 A; 13068 C; 14206 G; 19620 T; 0 U; 0 Other;

  Query Match             10.1%;  Score 202.4;  DB 14;  Length 64829;
  Best Local Similarity   77.7%;  Pred. No. 1.9e-40;
  Matches  283;  Conservative    0;  Mismatches   76;  Indels    5;  Gaps    3;

Qy       1192 AGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCA 1251
             | | | || || | |||| |  | |  |||| | ||| ||||||||   | | ||||||
Db      36622 ATTTTCTGTTTTTTTTTTGTTTGTTTGTTTTTTGTTTTTTTTTTTTTTGAGACAGTCTCG 36563

Qy       1252 CTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTC 1311
             || |||| |||||||||||||| ||||||| |   ||| ||||||||||||||||||||||
Db      36562 CTCTGTCGCCCAGGCTGGAGTGCAGTGGC-GCAATCTTGGCTCACTGCAACCTCTGCCTC 36504

Qy       1312 CCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC 1369
             ||  |||||||||||||||||  ||||| |||| | ||||||||||||||||||    |
Db      36503 CCGAGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCCCAC 36444

Qy       1370 ACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAG 1427
                | | |  |||||||||||||||||||||||||||| ||||||||||||| | ||
Db      36443 CATCACGTTCAGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCGCCATGTTGGCTAG 36384

Qy       1428 GCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGG 1487
             |||    || ||||| ||||| |||||||||||||||||||||||||||||  | ||
Db      36383 GCTGCTCTCGAACTTCTGACCTCAGGTGATCCACCCACCTCAGCCTCCCAAAGTGCAGGT 36324

Qy       1488 ATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAA 1547
             ||||||||| | |||||| | ||||||||| |||  ||  || |  ||  | ||  || |
Db      36323 ATTACAGGCGTTAGCCACTGCGCCCAGCCTCTTTGTTCTTATACATTAAATTTTCATGGA 36264

Qy       1548 GCCT 1551
                | |
Db      36263 CCGT 36260


RESULT 31
AAS36758
ID   AAS36758 standard; DNA; 12149 BP.
XX
AC   AAS36758;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human cardiovascular system antigen genomic DNA SEQ ID No 2258.
XX
KW   Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW   chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; tissue regeneration;
KW   anti-infertility.
XX
OS   Homo sapiens.
XX
PN   WO200155321-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001340.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
```

```
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451930/48.
XX
PT   New cardiovascular system related polynucleotides and polypeptides,
PT   useful for diagnosing, treating and/or preventing disorders of the
PT   cardiovascular system.
XX
PS   Claim 1; SEQ ID NO 2258; 674pp; English.
XX
CC   Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC   the cardiovascular system antigen polypeptides of the invention.
CC   Cardiovascular system antigens and their associated polynucleotides are
CC   useful in the diagnosis, treatment and prevention of various types of
CC   disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC   chickens or sheep. A pathological condition can be determined by
CC   detecting the presence or absence of a mutation in a cardiovascular
CC   system antigen polynucleotide. The treatable disorders include autoimmune
CC   diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC   as neoplasms of the breast or liver, cardiovascular disorders such as
CC   cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC   nervous system disorders such as Alzheimer's disease, infections caused
CC   by bacteria, viruses and fungi, ocular disorders such as corneal
CC   infection, endocrine disorders such as premature labour and infertility,
CC   gastrointestinal disorders such as Crohn's disease, renal disorders such
CC   as glomerulonephritis and respiratory disorders such as asthma and
CC   pleurisy. The polypeptides can also be used to aid wound healing, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, to regenerate tissues and in chemotaxis. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX

  Query Match             10.1%;  Score 202.2;  DB 4;  Length 12149;
  Best Local Similarity   76.4%;  Pred. No. 1.1e-40;
  Matches  288;  Conservative    0;  Mismatches   83;  Indels    6;  Gaps    3;

Qy       1196 TTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTT 1255
             |||  || ||||||      | ||||| ||||  | |||||||| |||||||||||||| |
Db       9764 TTTTGTTTTGTTTTGGTTTGTTTTTTTGTTTTGTTGTTTTTTGAGACAGAGTCTCACTCT 9823

Qy       1256 GTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAG 1315
             |||||||||||||||||| |||||| || | |||||||||||||||||||| ||||||  |
Db       9824 GTCACCCAGGCTGGAGTGCAGTGG-TGCGATCTCGGCTCACTGCAACCTCCGCCTCCTGG 9882

Qy       1316 GTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG---CA 1370
             ||||||  |||||||||| .||||| ||  || |||||||||||||||||||| |   |
Db       9883 GTTCAACGGATTCTCCTGCCTCAGCCTCTGGAGTAGCTGGGATTACAGGCGCACGCCACC 9942

Qy       1371 CCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
              | ||  ||  | ||||||||||||||||||||  | |||||||||||||| |||||||
Db       9943 ACGCCTGGCTAATTTTTTGTATTTTTAGTAGAGATGGGGTTTCGCCATGTTGGCCAGGCT 10002

Qy       1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490
              | ||  |||||||| | |||||||||| || |||||||||||||||  ||||||||
Db      10003 GGTCTCAAACTCCTGGCCTCAGGTGATCCGCCTACCTCAGCCTCCCAAAGTGCTGGGATT 10062

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCC 1550
             |||||||||||||||| | |   || ||  | ||   |||| |  ||     |
Db      10063 ACAGGCATGAGCCACCGCGTCTGTCCAGTCCTTTCCTTTCCTTTCCTTTCCATTTTCCTT 10122

Qy       1551 TTCATTTCTATCTTCTT 1567
              | | || || || | ||
Db      10123 TCCTTTCCTTTCCTTTT 10139


RESULT 32
ADE47452
ID   ADE47452 standard; DNA; 12149 BP.
XX
AC   ADE47452;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human cardiovascular system related genomic DNA #1018.
XX
KW   Human; cardiovascular system related polypeptide; cancer;
KW   proliferative disorder; foetal abnormality; developmental abnormality;
KW   haematopoietic disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; gene; ds.
XX
OS   Homo sapiens.
```

XX
PN US2003059908-A1.
XX
PD 27-MAR-2003.
XX
PF 07-MAR-2002; 2002US-00091504.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764869.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

```
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-743766/70.
XX
PT   New cardiovascular system related polynucleotides and polypeptides,
PT   useful for preventing, treating, or ameliorating a medical condition,
PT   such as cancer of cardiovascular tissues and cancer metastases.
XX
PS   Claim 1; SEQ ID NO 2258; 262pp; English.
XX
CC   The invention relates to human cardiovascular system related polypeptides
CC   and the polynucleotides encoding them. The polypeptides, polynucleotides
CC   and antibodies to the polypeptides are useful for diagnosing a
CC   pathological condition or a susceptibility to a pathological condition,
CC   for preventing, treating, or ameliorating a medical condition, such as
CC   cancer of cardiovascular system tissues, proliferative disorders, foetal
CC   and developmental abnormalities, haematopoietic disorders, diseases of
CC   the immune system, AIDS, autoimmune diseases (e.g., rheumatoid
CC   arthritis), inflammation, allergies, neurological disorders (e.g.,
CC   Alzheimer's disease, Parkinson's disease), cognitive disorders,
CC   schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis,
CC   diabetes, atherosclerosis, cardiovascular disorders, angiogenic
CC   disorders, kidney disorders, gastrointestinal disorders, pregnancy-
CC   related disorders, endocrine disorders and infections. The nucleic acids
CC   are also useful for chromosome identification, radiation hybrid mapping
CC   or long-range restriction mapping. The polypeptides and polynucleotides
CC   may also be used as food additives or preservatives to increase or
CC   decrease storage capabilities, fat content or other nutritional
CC   components. This sequence represents human cardiovascular system related
CC   genomic DNA of the invention.
XX
SQ   Sequence 12149 BP; 2788 A; 3327 C; 3233 G; 2801 T; 0 U; 0 Other;

  Query Match             10.1%;  Score 202.2;  DB 10;  Length 12149;
  Best Local Similarity   76.4%;  Pred. No. 1.1e-40;
  Matches  288;  Conservative    0;  Mismatches   83;  Indels    6;  Gaps    3;

Qy       1196 TTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTT 1255
             |||  || ||||||      | ||||| ||||  | |||||||| ||||||||||||| |
Db       9764 TTTTGTTTTGTTTTGGTTTGTTTTTTTGTTTTGTTGTTTTTTGAGACAGAGTCTCACTCT 9823

Qy       1256 GTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAG 1315
             |||||||||||||||| |||||| || | |||||||||||||||||||| |||||| |
Db       9824 GTCACCCAGGCTGGAGTGCAGTGG-TGCGATCTCGGCTCACTGCAACCTCCGCCTCCTGG 9882

Qy       1316 GTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG---CA 1370
             ||||||  |||||||||||  ||||| ||  || |||||||||||||||||||| |   |
Db       9883 GTTCAACGGATTCTCCTGCCTCAGCCTCTGGAGTAGCTGGGATTACAGGCGCACGCCACC 9942

Qy       1371 CCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
               | ||  ||  | |||||||||||||||||||||  | |||||||||||||| |||||||
Db       9943 ACGCCTGGCTAATTTTTTGTATTTTTAGTAGAGATGGGGTTTCGCCATGTTGGCCAGGCT 10002

Qy       1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490
               | ||  |||||||| | |||||||||| || ||||||||||||||||  ||||||||
Db      10003 GGTCTCAAACTCCTGGCCTCAGGTGATCCGCCTACCTCAGCCTCCCAAAGTGCTGGGATT 10062

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCC 1550
             |||||||||||||||| | |   || ||  | ||   |||| |  ||     |
Db      10063 ACAGGCATGAGCCACCGCGTCTGTCCAGTCCTTTCCTTTCCTTTCCTTTCCATTTTCCTT 10122

Qy       1551 TTCATTTCTATCTTCTT 1567
             | | || || || | ||
Db      10123 TCCTTTCCTTTCCTTTT 10139


RESULT 33
ADJ08870
ID   ADJ08870 standard; DNA; 12149 BP.
XX
AC   ADJ08870;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Human cardiovascular system associated polypeptide-related DNA SeqID2258.
XX
KW   autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW   breast neoplasms; liver neoplasm; cardiovascular disorder;
KW   cardiac arrest; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   ocular disorder; corneal infection; wound healing;
KW   epithelial cell proliferation; skin aging; sunburn;
KW   organ transplantation; cell culture; tissue regeneration; chemotaxis;
KW   food additive; preservative; cardiovascular system associated antigen;
KW   nuclear factor kappaB; NFkappaB; promoter element; human; ds.
XX
OS   Homo sapiens.
XX
PN   US2004005575-A1.
XX
PD   08-JAN-2004.
XX
PF   26-AUG-2002; 2002US-00227577.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
```

PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764869.
PR 07-MAR-2002; 2002US-00091504.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2004-081713/08.
XX
PT New cardiovascular system-related nucleic acid molecule, useful for
PT diagnosing, preventing or treating diseases of the cardiovascular system,
PT and in chromosome mapping, drug screening or in pharmacogenomics.
XX
PS Disclosure; SEQ ID NO 2258; 262pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule encoding a
CC human cardiovascular system associated polypeptide (or antigens), or its
CC fragment. Also included recombinant vectors, recombinant host cells, an
CC isolated human cardiovascular system associated polypeptide (including
CC its fragment, allelic variant, species homologue or epitope), an isolated
CC antibody that binds specifically to a human cardiovascular system
CC associated polypeptide, diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or absence of a mutation in human cardiovascular system
CC associated nucleic acid and diagnosing a condition based on the presence
CC or absence of the mutation), identifying a binding partner to human
CC cardiovascular system associated polypeptides, the gene corresponding to
CC the human cardiovascular system associated cDNA sequence and identifying
CC an activity in a biological assay comprising expressing the human
CC cardiovascular system associated cDNA in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity. The human
CC cardiovascular system associated nucleic acids and polypeptides are used
CC to prevent, treat or ameliorate a medical condition (for example in
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep), for
CC example autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders, for example neoplasms of the breast or
CC liver, cardiovascular disorders, for example cardiac arrest,

Query Match 10.1%; Score 202.2; DB 13; Length 12149;
Best Local Similarity 76.4%; Pred. No. 1.1e-40;
Matches 288; Conservative 0; Mismatches 83; Indels 6; Gaps 3;

Qy 1196 TTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTT 1255
Db 9764 TTTTGTTTTGTTTTGGTTTGTTTTTTTTGTTTTGTTGTTTTTTGAGACAGAGTCTCACTCT 9823

Qy 1256 GTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAG 1315
Db 9824 GTCACCCAGGCTGGAGTGCAGTGG-TGCGATCTCGGCTCACTGCAACCTCCGCCTCCTGG 9882

Qy 1316 GTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG---CA 1370
Db 9883 GTTCAACGGATTCTCCTGCCTCAGCCTCTGGAGTAGCTGGGATTACAGGCGCACGCCACC 9942

Qy 1371 CCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
Db 9943 ACGCCTGGCTAATTTTTTGTATTTTTAGTAGAGATGGGGTTTCGCCATGTTGGCCAGGCT 10002

Qy 1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490

```
Db      10003 GGTCTCAAACTCCTGGCCTCAGGTGATCCGCCTACCTCAGCCTCCCAAAGTGCTGGGATT 10062

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCC 1550
              |||||||||||||||||| | |   || ||  | ||   |||| |  ||     |
Db      10063 ACAGGCATGAGCCACCGCGTCTGTCCAGTCCTTTCCTTTCCTTTCCTTTCCATTTTCCTT 10122

Qy       1551 TTCATTTCTATCTTCTT 1567
              | | || || || | ||
Db      10123 TCCTTTCCTTTCCTTTT 10139


RESULT 34
AEB32377/c
ID   AEB32377 standard; DNA; 160274 BP.
XX
AC   AEB32377;
XX
DT   08-SEP-2005  (first entry)
XX
DE   Human genomic DNA #18.
XX
KW   SNP detection; diagnosis; non-insulin dependent diabetes; obesity;
KW   antidiabetic; anorectic; endocrine disease; gastrointestinal disease;
KW   metabolic disorder; nutritional disorder; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2005147987-A1.
XX
PD   07-JUL-2005.
XX
PF   19-JUL-2004; 2004US-00893315.
XX
PR   08-SEP-2000; 2000US-0231397P.
PR   10-SEP-2001; 2001US-00948947.
XX
PA   (APPL-) APPLERA CORP NY.
XX
PI   Venter JC,  Zhang JN,  Liu X,  Rowe W,  Cravchik A,  Kalush F;
PI   Naik A,  Subramanian G,  Woodage T;
XX
DR   WPI; 2005-511776/52.
XX
PT   New detection reagent capable of detecting 1, 100, 500, 1000 or 5000 or
PT   more single nucleic acid polymorphisms, useful in identifying an
PT   individual having or at risk of developing type II diabetes or obesity.
XX
PS   Disclosure; SEQ ID NO 140; 31pp; English.
XX
CC   The invention relates to a detection reagent capable of detecting one or
CC   more single nucleic acid polymorphisms. The invention also relates to
CC   determining whether a trait is linked to one of the human chromosomes or
CC   its sub-region, a computer readable medium having stored in it the SNP
CC   relational information given in the specification, an isolated nucleic
CC   acid molecule for detecting at least one SNP given in the specification
CC   comprising at least about 12 contiguous nucleotides, genotyping at least
CC   one SNP position given in the specification in a sample, identifying an
CC   individual having or at risk of developing a disorder and a kit
CC   comprising at least one container containing the detection reagent.
CC   Determining whether a trait is linked to one of the human chromosomes or
CC   its sub-region comprises determining whether the trait is linked to one
CC   or more SNPs using the detection reagents. Genotyping at least one SNP
CC   position given in the specification in a sample comprises contacting the
CC   sample with a detection reagent that differentiates between alternative
CC   alleles at at least one SNP position given in the specification, and
CC   determining which allele is present at the at least one SNP position.
CC   Identifying an individual having or at risk of developing a disorder
CC   comprises genotyping at least one SNP given in the specification in a
CC   nucleic acid sample from the individual. The disorder is type II diabetes
CC   (non-insulin dependent diabetes) or obesity. The detection reagent is
CC   useful in identifying an individual having or at risk of developing a
CC   disorder, particularly type II diabetes or obesity. This sequence
CC   represents human genomic DNA used in the scope of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 160274 BP; 50545 A; 28535 C; 28704 G; 52490 T; 0 U; 0 Other;

  Query Match           10.1%;  Score 202;  DB 14;  Length 160274;
  Best Local Similarity   77.2%;  Pred. No. 3.5e-40;
  Matches  284;  Conservative    0;  Mismatches   80;  Indels    4;  Gaps    3;

Qy       1199 GATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTC 1258
              |||  | |||  |   | | ||||||||  |||||||||| |||| |||||||| ||||
Db      21584 GATAGTCTTTATTTTATTTTATTTATTTTATTTTTTTTTGAGACAGGGTCTCACTCTGTC 21525

Qy       1259 ACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTT 1318
              ||||||| ||| || |||||| ||   |||| ||||||||||||||||  ||||||||||
Db      21524 ACCCAGGTTGGCGTGCAGTGG-TGCAATCTCAGCTCACTGCAACCTCCACCTCCCAGGTT 21466

Qy       1319 CAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC-C 1375
              ||||| |||||||||  ||||| |||| | |||||||| |||||| |    ||||||  |
Db      21465 CAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGTGGTGCCACCACAC 21406

Qy       1376 ATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCT 1435
                | |||||||||||||||||||||||| | ||||| |||||||| ||| ||| | ||
Db      21405 TAGGCTAATTTTTGTATTTTTAGTAGAGACTGGGTTTCACCATGTTGGCCAAGCTGGTCT 21346

Qy       1436 TGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGG 1495
              |||||||||||| ||| |||||||||| |||||||| ||||||  ||| |||||||||
Db      21345 TGAACTCCTGACCTCAAGTGATCCACCCGCCTCAGCCTTCCAAAGTGCTGAGATTACAGG 21286

Qy       1496 CATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCAT 1555
              |||| ||||| |||| ||||| || |       |     ||||| | | | ||   || ||
Db      21285 TATGATCCACCATGCCTAGCCTATTCTTAATTTTTTAATATTTTTATTTCAACAGTTAAT 21226

Qy       1556 TTCTATCT 1563
              |    |||
Db      21225 TAGAGTCT 21218


RESULT 35
AEB32388/c
ID   AEB32388 standard; DNA; 160300 BP.
XX
AC   AEB32388;
XX
DT   08-SEP-2005  (first entry)
XX
DE   Human genomic DNA #29.
XX
KW   SNP detection; diagnosis; non-insulin dependent diabetes; obesity;
KW   antidiabetic; anorectic; endocrine disease; gastrointestinal disease;
KW   metabolic disorder; nutritional disorder; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2005147987-A1.
XX
PD   07-JUL-2005.
XX
PF   19-JUL-2004; 2004US-00893315.
XX
PR   08-SEP-2000; 2000US-0231397P.
PR   10-SEP-2001; 2001US-00948947.
XX
PA   (APPL-) APPLERA CORP NY.
XX
PI   Venter JC,  Zhang JN,  Liu X,  Rowe W,  Cravchik A,  Kalush F;
PI   Naik A,  Subramanian G,  Woodage T;
XX
DR   WPI; 2005-511776/52.
XX
PT   New detection reagent capable of detecting 1, 100, 500, 1000 or 5000 or
PT   more single nucleic acid polymorphisms, useful in identifying an
```

```
PT   individual having or at risk of developing type II diabetes or obesity.
XX
PS   Disclosure; SEQ ID NO 151; 31pp; English.
XX
CC   The invention relates to a detection reagent capable of detecting one or
CC   more single nucleic acid polymorphisms. The invention also relates to
CC   determining whether a trait is linked to one of the human chromosomes or
CC   its sub-region, a computer readable medium having stored in it the SNP
CC   relational information given in the specification, an isolated nucleic
CC   acid molecule for detecting at least one SNP given in the specification
CC   comprising at least about 12 contiguous nucleotides, genotyping at least
CC   one SNP position given in the specification in a sample, identifying an
CC   individual having or at risk of developing a disorder and a kit
CC   comprising at least one container containing the detection reagent.
CC   Determining whether a trait is linked to one of the human chromosomes or
CC   its sub-region comprises determining whether the trait is linked to one
CC   or more SNPs using the detection reagents. Genotyping at least one SNP
CC   position given in the specification in a sample comprises contacting the
CC   sample with a detection reagent that differentiates between alternative
CC   alleles at at least one SNP position given in the specification, and
CC   determining which allele is present at the at least one SNP position.
CC   Identifying an individual having or at risk of developing a disorder
CC   comprises genotyping at least one SNP given in the specification in a
CC   nucleic acid sample from the individual. The disorder is type II diabetes
CC   (non-insulin dependent diabetes) or obesity. The detection reagent is
CC   useful in identifying an individual having or at risk of developing a
CC   disorder, particularly type II diabetes or obesity. This sequence
CC   represents human genomic DNA used in the scope of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 160300 BP; 50550 A; 28538 C; 28712 G; 52500 T; 0 U; 0 Other;

  Query Match             10.1%;  Score 202;  DB 14;  Length 160300;
  Best Local Similarity   77.2%;  Pred. No. 3.5e-40;
  Matches  284;  Conservative    0;  Mismatches   80;  Indels    4;  Gaps    3;

Qy       1199 GATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCACTTTGTC 1258
             |||   | |||   |     | | ||||||||  |||||||||| |||| |||||||| ||||
Db      21586 GATAGTCTTTATTTTATTTTTATTTATTTTTATTTTTTTTTTGAGACAGGGTCTCACTCTGTC 21527

Qy       1259 ACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTT 1318
             ||||||| ||| || |||||| ||    |||| ||||||||||||||  ||||||||||
Db      21526 ACCCAGGTTGGCGTGCAGTGG-TGCAATCTCAGCTCACTGCAACCTCCACCTCCCAGGTT 21468

Qy       1319 CAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC-C 1375
             |||||  |||||||||  |||||  |||| | |||||||||  ||||||  |    |||||| |
Db      21467 CAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGTGGTGCCACCACAC 21408

Qy       1376 ATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCT 1435
               | ||||||||||||||||||||||||| | |||||  |||||||  |||  |||  | ||
Db      21407 TAGGCTAATTTTTGTATTTTTAGTAGAGACTGGGTTTCACCATGTTGGCCAAGCTGGTCT 21348

Qy       1436 TGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGG 1495
             ||||||||||||  |||  ||||||||||| ||||||||| ||||||   |||  |||||||||
Db      21347 TGAACTCCTGACCTCAAGTGATCCACCCGCCTCAGCCTTCCAAAGTGCTGAGATTACAGG 21288

Qy       1496 CATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCAT 1555
               ||||  |||||  ||||  |||||  || |       |    |||||  | | | ||   || ||
Db      21287 TATGATCCACCATGCCTAGCCTATTCTTAATTTTTTAATATTTTTATTTCAACAGTTAAT 21228

Qy       1556 TTCTATCT 1563
             |    |||
Db      21227 TAGAGTCT 21220


RESULT 36
ADB85543/c
ID   ADB85543 standard; DNA; 51256 BP.
XX
AC   ADB85543;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human sentrin-specific family-related protease genomic DNA sequence.
XX
KW   human protease protein; sentrin-specific protease subfamily; SENP;
KW   sentrinisation; acute myelocytic leukaemia; placenta; breast; testis;
KW   head; neck; cervix; melanocytes; uterus;
KW   high grade serous papillary carcinoma tumour; marrow; human; gene; ds;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(425,A)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1521. .1522,T)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(2245,A)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(2492,G)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             3000. .49059
FT                   /*tag=  f
FT   exon            3000. .4512
FT                   /*tag=  e
FT                   /number= 1
FT   variation       replace(4019,C)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4048,A)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          4513. .17481
FT                   /*tag=  i
FT                   /number= 1
FT   variation       replace(5319,A)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5711,G)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5886,C)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5895,A)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6392,T)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7540,C)
FT                   /*tag=  o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7638,A)
FT                   /*tag=  p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8154,T)
FT                   /*tag=  q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8325,C)
FT                   /*tag=  r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8344,G)
FT                   /*tag=  s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8378. .8379,A)
FT                   /*tag=  t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8430,T)
```

```
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8986,C)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9144. .9145,TA, T)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9263,T)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9589,C)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10421. .10422,AT, A)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11453,T)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12362. .12363,T)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13896,C)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14287. .14288,T)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14725,G)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14947,C)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15600,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15764,A)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16684,G)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16913,T, C)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17088,C)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17220. .17221,C)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17293,A)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            17482. .17587
FT                   /*tag= an
FT                   /number= 2
FT   intron          17588. .17739
FT                   /*tag= ao
FT                   /number= 2
FT   variation       replace(17634,A)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            17740. .17878
FT                   /*tag= aq
FT                   /number= 3
FT   intron          17879. .18180
FT                   /*tag= ar
FT                   /number= 3
FT   variation       replace(18033,G)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            18181. .18228
FT                   /*tag= at
FT                   /number= 4
FT   intron          18229. .21348
FT                   /*tag= au
FT                   /number= 4
FT   variation       replace(18280,C)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18301,T)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18306,A)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18458,A)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20037,G)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            21349. .21426
FT                   /*tag= ba
FT                   /number= 5
FT   intron          21427. .41548
FT                   /*tag= bb
FT                   /number= 5
FT   variation       replace(21591,A)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22083,A)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22258,G)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23328,T)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25434. .25435,AA, A)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25869,G)
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25898,G)
FT                   /*tag= bi
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26048,C)
```

```
  Query Match             10.1%;  Score 201.6;  DB 10;  Length 51256;
  Best Local Similarity   82.0%;  Pred. No. 2.8e-40;
  Matches  269;  Conservative    0;  Mismatches   54;  Indels    5;  Gaps    3;

Qy       1221 TTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGC 1280
              || ||||| |||||||||| |||||||||||||| |||||||||||||||||| ||||||
Db      11744 TTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGG- 11686

Qy       1281 TGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGC 1338
              || | ||||  |||||||||||||||   ||| || || ||||| |||||||||  |||||
Db      11685 TGCGATCTCAACTCACTGCAACCTCCAACTCACAAGTACAAGCAATTCTCCTGTCTCAGC 11626

Qy       1339 TTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTATTTTT 1396
              |||| | |||||||| ||||||||||||||    |    | || ||||||||||||||||
Db      11625 CTCCCAAGTAGCTGGGACTACAGGCGCATGCCACCATGCCTGGCTAATTTTTGTATTTTT 11566

Qy       1397 AGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGA 1456
              ||||||||| ||||||| |||| |||  |||||| | |||||||||||||| ||||||||
Db      11565 AGTAGAGACGGAGTTTCACCATATTGGTCAGGCTAGTCTTGAACTCCTGACCTCAGGTGA 11506

Qy       1457 TCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCC 1516
              |||||||||||| |||||||||||  ||||||||||||||||||||||||| |||| |||
```

```
Db      11505 TCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGGCC 11446

Qy       1517 TGTTTTCTCAGATCCTGTATTTGTTTCT 1544
              |    |  |   || || |  || || |
Db      11445 TAAACTTACTCTTCTTGAAGCTGCTTTT 11418


RESULT 37
ADK66161/c
ID   ADK66161 standard; DNA; 51256 BP.
XX
AC   ADK66161;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human protease gene.
XX
KW   Protease; therapeutic; drug; diagnosis; pharmacogenomic analysis;
KW   treatment; human; chromosome 3; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(425,G)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1521. .1522,TG)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(2245,A)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(2492,G)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   CDS             3000. .49059
FT                   /*tag=  f
FT                   /product= "Human protease protein"
FT   exon            3000. .4512
FT                   /*tag=  e
FT   variation       replace(4019,C)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(4048,A)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   intron          4513. .17481
FT                   /*tag=  i
FT   variation       replace(5319,A)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(5711,G)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(5886,C)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(5895,A)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(6392,T)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(7540,C)
FT                   /*tag=  o
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(7638,A)
FT                   /*tag=  p
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(8154,T)
FT                   /*tag=  q
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(8325,C)
FT                   /*tag=  r
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(8344,G)
FT                   /*tag=  s
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(8378. .8380,AT)
FT                   /*tag=  t
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(8430,T)
FT                   /*tag=  u
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(8986,C)
FT                   /*tag=  v
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9144. .9146,AT)
FT                   /*tag=  w
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9145,A)
FT                   /*tag=  x
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9263,T)
FT                   /*tag=  y
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9589,C)
FT                   /*tag=  z
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10421. .10423,AA)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10422,T)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(11453,T)
FT                   /*tag=  ac
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12362. .12364,TA)
FT                   /*tag=  ad
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13896,C)
FT                   /*tag=  ae
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14287. .14289,TG)
FT                   /*tag=  af
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14725,G)
FT                   /*tag=  ag
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14947,C)
FT                   /*tag=  ah
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15600,G)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15764,A)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16684,G)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16913,C, T)
FT                   /*tag=  al
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17088,C)
FT                   /*tag=  am
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17220. .17222,CA)
FT                   /*tag=  an
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17293,A)
FT                   /*tag=  ao
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            17482. .17587
FT                   /*tag=  ap
```

```
FT   intron          17588. .17739
FT                   /*tag=  aq
FT   variation       replace(17634,A)
FT                   /*tag=  ar
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            17740. .17878
FT                   /*tag=  as
FT   intron          17879. .18180
FT                   /*tag=  at
FT   variation       replace(18033,G)
FT                   /*tag=  au
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            18181. .18228
FT                   /*tag=  av
FT   intron          18229. .21348
FT                   /*tag=  aw
FT   variation       replace(18280,C)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18301,T)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18306,A)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18458,A)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(20037,G)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            21349. .21426
FT                   /*tag=  bc
FT   intron          21427. .41548
FT                   /*tag=  bd
FT   variation       replace(21591,A)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22083,A)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22258,G)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(23228,T)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(25434. .25436,AA)
FT                   /*tag=  bi
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(25435,A)
FT                   /*tag=  bj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(25869,G)
FT                   /*tag=  bk
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(25898,G)
FT                   /*tag=  bl
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(26048,C)
FT                   /*tag=  bm
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(26191,A)

  Query Match             10.1%;  Score 201.6;  DB 10;  Length 51256;
  Best Local Similarity   82.0%;  Pred. No. 2.8e-40;
  Matches  269;  Conservative    0;  Mismatches   54;  Indels    5;  Gaps    3;

Qy       1221 TTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGC 1280
              || |||||  |||||||||| |||||||||||||| |||||||||||||||||||| ||||||
Db      11744 TTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGG- 11686

Qy       1281 TGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGC 1338
              || | ||||  |||||||||||||||   ||| || || ||||| |||||||||  |||||
Db      11685 TGCGATCTCAACTCACTGCAACCTCCAACTCACAAGTACAAGCAATTCTCCTGTCTCAGC 11626

Qy       1339 TTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTATTTTT 1396
              |||| | ||||||||| ||||||||||||||    |    | || ||||||||||||||||||
Db      11625 CTCCCAAGTAGCTGGGACTACAGGCGCATGCCACCATGCCTGGCTAATTTTTGTATTTTT 11566

Qy       1397 AGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGA 1456
              ||||||||| ||||||| |||| |||  |||||| | |||||||||||||| ||||||||
Db      11565 AGTAGAGACGGAGTTTCACCATATTGGTCAGGCTAGTCTTGAACTCCTGACCTCAGGTGA 11506

Qy       1457 TCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCC 1516
              |||||||||||| |||||||||||  ||||||||||||||||||||||||| |||| |||
Db      11505 TCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGGCC 11446

Qy       1517 TGTTTTCTCAGATCCTGTATTTGTTTCT 1544
              |    |  |   || || |  || || |
Db      11445 TAAACTTACTCTTCTTGAAGCTGCTTTT 11418
```

```
RESULT 38
ABD33611/c
ID   ABD33611 standard; DNA; 89567 BP.
XX
AC   ABD33611;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-124.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 838; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 89567 BP; 24453 A; 17570 C; 19182 G; 28362 T; 0 U; 0 Other;
```

```
  Query Match             10.1%;  Score 201.6;  DB 13;  Length 89567;
  Best Local Similarity   84.3%;  Pred. No. 3.5e-40;
  Matches  263;  Conservative    0;  Mismatches   44;  Indels    5;  Gaps    3;

Qy       1217 CTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAG 1276
             || ||| || || |||||||||| |||||||||||| |||| ||||||||||||||||||
Db      37961 CTGTTTTTTGTTTTTTTTTTTGAGACAGAGTCTCACTGTGTCTCCCAGGCTGGAGTACAG 37902

Qy       1277 TGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--T 1334
             ||| ||   |||||| ||||||||||||||||||  ||||||||||||||||||   |
Db      37901 TGG-TGCAATCTCGGTTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTCCTGCCT 37843

Qy       1335 CAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTAT 1392
             |||| ||| || ||||||||||||||| ||||||   |   |    ||||||||||||
Db      37842 CAGCCTCCTGAGTAGCTGGGATTACAGGTGCATGCCATCATGCCCAGCTAATTTTTGTAT 37783

Qy       1393 TTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAG 1452
             |||||||||||||| | ||||| |||||||| ||||||| | ||  ||||||||||  |||
Db      37782 TTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCCCAG 37723

Qy       1453 GTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCC 1512
              |||||||||||||||||||||||||  |||||||||||||| |||||||| |||||
Db      37722 ATGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCATGCCC 37663

Qy       1513 AGCCTGTTTTCT 1524
              |||||  | | |
Db      37662 AGCCTCCTGTGT 37651


RESULT 39
ADE43582
ID   ADE43582 standard; DNA; 128034 BP.
XX
AC   ADE43582;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Polymorphic human IDE genomic sequence, SEQ ID 187.
XX
KW   Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;
KW   Alzheimer's disease; neuroprotective; nootropic; gene therapy;
KW   Chromosome 10; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    2456
FT                   /*tag=  a
FT                   /note= "There is a variation at this position"
FT   misc_feature    3279
FT                   /*tag=  b
FT                   /note= "There is a variation at this position"
FT   misc_feature    3407
FT                   /*tag=  c
FT                   /note= "There is a variation at this position"
FT   misc_feature    17095
FT                   /*tag=  d
FT                   /note= "There is a variation at this position"
FT   misc_feature    17242
FT                   /*tag=  e
FT                   /note= "There is a variation at this position"
FT   misc_feature    33590
FT                   /*tag=  f
FT                   /note= "There is a variation at this position"
FT   misc_feature    38903
FT                   /*tag=  g
FT                   /note= "There is a variation at this position"
FT   misc_feature    42943
FT                   /*tag=  h
FT                   /note= "There is a variation at this position"
FT   misc_feature    43391
FT                   /*tag=  i
FT                   /note= "There is a variation at this position"
FT   misc_feature    45017
FT                   /*tag=  j
FT                   /note= "There is a variation at this position"
FT   misc_feature    62498
FT                   /*tag=  k
FT                   /note= "There is a variation at this position"
FT   misc_feature    68906
FT                   /*tag=  l
FT                   /note= "There is a variation at this position"
FT   misc_feature    68973
FT                   /*tag=  m
FT                   /note= "There is a variation at this position"
FT   misc_feature    69586
FT                   /*tag=  n
FT                   /note= "There is a variation at this position"
FT   misc_feature    73772
FT                   /*tag=  o
FT                   /note= "There is a variation at this position"
FT   misc_feature    74084
FT                   /*tag=  p
FT                   /note= "There is a variation at this position"
FT   misc_feature    83024
FT                   /*tag=  q
FT                   /note= "There is a variation at this position"
FT   misc_feature    83104
FT                   /*tag=  r
FT                   /note= "There is a variation at this position"
FT   misc_feature    89301
FT                   /*tag=  s
FT                   /note= "There is a variation at this position"
FT   misc_feature    105060
FT                   /*tag=  t
FT                   /note= "There is a variation at this position"
FT   misc_feature    107395
FT                   /*tag=  u
FT                   /note= "There is a variation at this position"
FT   misc_feature    108489
FT                   /*tag=  v
FT                   /note= "There is a variation at this position"
FT   misc_feature    111914
FT                   /*tag=  w
FT                   /note= "There is a variation at this position"
FT   misc_feature    112114
FT                   /*tag=  x
FT                   /note= "There is a variation at this position"
FT   misc_feature    113142
FT                   /*tag=  y
FT                   /note= "There is a variation at this position"
FT   misc_feature    113591
FT                   /*tag=  z
FT                   /note= "There is a variation at this position"
FT   misc_feature    114683
FT                   /*tag=  aa
FT                   /note= "There is a variation at this position"
FT   misc_feature    116662
FT                   /*tag=  ab
FT                   /note= "There is a variation at this position"
FT   misc_feature    117803
FT                   /*tag=  ac
FT                   /note= "There is a variation at this position"
FT   misc_feature    124565
FT                   /*tag=  ad
FT                   /note= "There is a variation at this position"
XX
PN   WO2003054143-A2.
XX
PD   03-JUL-2003.
XX
PF   25-OCT-2002; 2002WO-US034679.
XX
PR   25-OCT-2001; 2001US-0339525P.
PR   08-NOV-2001; 2001US-0336929P.
PR   08-NOV-2001; 2001US-0338010P.
```

```
PR   09-NOV-2001; 2001US-0338363P.
PR   04-DEC-2001; 2001US-0337052P.
PR   28-MAR-2002; 2002US-0368919P.
XX
PA   (NEUR-) NEUROGENETICS INC.
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Becker KD, Velicelebi G, Elliott KJ, Wang X, Tanzi RE, Bertram L;
PI   Saunders AJ, Mullin KM, Sampson AJ, Blacker DL;
XX
DR   WPI; 2003-559131/52.
XX
PT   Determining a predisposition for or the occurrence of neurodegenerative
PT   disease, e.g. Alzheimer's disease by detecting in a target nucleic acid
PT   the presence or absence of an allelic variant of one or more polymorphic
PT   regions.
XX
PS   Claim 9; Page 584-page 618; 848pp; English.
XX
CC   The present invention relates to a method (M1) for determining a
CC   predisposition for or the occurrence of neurodegenerative disease in a
CC   subject. The method comprises detecting in a target nucleic acid obtained
CC   from the subject the presence or absence of an allelic variant of one or
CC   more polymorphic regions of one or more genes selected from uPA
CC   (Urokinase plasminogen activator), SNCG (gamma-synuclein), IDE (insulin-
CC   degrading enzyme), KNSL1 (Kinesin-like protein 1), LIPA (lysosomal acid
CC   lypase), and TNFRSF6 (Tumour Necrosis Factor Receptor-SF6), where the
CC   presence of at least one of the allelic variant of one or more
CC   polymorphic regions is indicative of a predisposition for or the
CC   occurrence of neurodegenerative disease. The genes are all located on
CC   chromosome 10. M1 is useful for determining a predisposition for or the
CC   occurrence of, and for treating neurodegenerative disease, particularly
CC   Alzheimer's disease.
XX
SQ   Sequence 128034 BP; 34726 A; 25977 C; 26400 G; 40799 T; 0 U; 132 Other;

  Query Match             10.1%;  Score 201.6;  DB 10;  Length 128034;
  Best Local Similarity   82.0%;  Pred. No. 4.1e-40;
  Matches  269;  Conservative    0;  Mismatches   54;  Indels    5;  Gaps    3;

Qy       1218 TTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGT 1277
             | ||  ||||| |||||||||| |||||||||||| |||  |||||||||||| ||||
Db      24034 TATTCTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAGT 24093

Qy       1278 GGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TC 1335
             |||  | |||| |||||||||||||||||||||||| |||||||| |||||||||  ||
Db      24094 GGC-ATAGTCTTGGCTCACTGCAACCTCTGCCTCCCAAGTTCAAGCAATTCTCCTGCCTC 24152

Qy       1336 AGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTATT 1393
             ||||||| || || |||||| |||||||  |||  |  ||  || ||||||||||||||
Db      24153 AGCTTCCTGAGTACCTGGGACTACAGGCGTGTGCCACCACACCTGGCTAATTTTTGTATT 24212

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             |||| |||||||||  |||| |||||||| |||  || | || ||||||||||||||||
Db      24213 TTTAATAGAGACAGGATTTCACCATGTTGGCCAACCTGGTCTAGAACTCCTGACTTCAGG 24272

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||| ||  |||||||||||||||  |||||||||||||  ||||||||| ||||||
Db      24273 TGATCCGCCTTCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCCA 24332

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             ||||| ||| ||   | ||   ||  ||
Db      24333 GCCTGCTTTTTCTTTTTCTTATTTATTT 24360


RESULT 40
ADE43581
ID   ADE43581 standard; DNA; 128034 BP.
XX
AC   ADE43581;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human IDE genomic sequence, SEQ ID 186.
XX
KW   Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;
KW   Alzheimer's disease; neuroprotective; nootropic; gene therapy;
KW   Chromosome 10; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003054143-A2.
XX
PD   03-JUL-2003.
XX
PF   25-OCT-2002; 2002WO-US034679.
XX
PR   25-OCT-2001; 2001US-0339525P.
PR   08-NOV-2001; 2001US-0336929P.
PR   08-NOV-2001; 2001US-0338010P.
PR   09-NOV-2001; 2001US-0338363P.
PR   04-DEC-2001; 2001US-0337052P.
PR   28-MAR-2002; 2002US-0368919P.
XX
PA   (NEUR-) NEUROGENETICS INC.
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Becker KD, Velicelebi G, Elliott KJ, Wang X, Tanzi RE, Bertram L;
PI   Saunders AJ, Mullin KM, Sampson AJ, Blacker DL;
XX
DR   WPI; 2003-559131/52.
XX
PT   Determining a predisposition for or the occurrence of neurodegenerative
PT   disease, e.g. Alzheimer's disease by detecting in a target nucleic acid
PT   the presence or absence of an allelic variant of one or more polymorphic
PT   regions.
XX
PS   Claim 22; Page 549-584; 848pp; English.
XX
CC   The present invention relates to a method (M1) for determining a
CC   predisposition for or the occurrence of neurodegenerative disease in a
CC   subject. The method comprises detecting in a target nucleic acid obtained
CC   from the subject the presence or absence of an allelic variant of one or
CC   more polymorphic regions of one or more genes selected from uPA
CC   (Urokinase plasminogen activator), SNCG (gamma-synuclein), IDE (insulin-
CC   degrading enzyme), KNSL1 (Kinesin-like protein 1), LIPA (lysosomal acid
CC   lypase), and TNFRSF6 (Tumour Necrosis Factor Receptor-SF6), where the
CC   presence of at least one of the allelic variant of one or more
CC   polymorphic regions is indicative of a predisposition for or the
CC   occurrence of neurodegenerative disease. The genes are all located on
CC   chromosome 10. M1 is useful for determining a predisposition for or the
CC   occurrence of, and for treating neurodegenerative disease, particularly
CC   Alzheimer's disease.
XX
SQ   Sequence 128034 BP; 34731 A; 25985 C; 26409 G; 40808 T; 0 U; 101 Other;

  Query Match             10.1%;  Score 201.6;  DB 10;  Length 128034;
  Best Local Similarity   82.0%;  Pred. No. 4.1e-40;
  Matches  269;  Conservative    0;  Mismatches   54;  Indels    5;  Gaps    3;

Qy       1218 TTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGT 1277
             | ||  ||||| |||||||||| |||||||||||| |||  |||||||||||| ||||
Db      24034 TATTCTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAGT 24093

Qy       1278 GGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TC 1335
             |||  | |||| |||||||||||||||||||||||| |||||||| |||||||||  ||
Db      24094 GGC-ATAGTCTTGGCTCACTGCAACCTCTGCCTCCCAAGTTCAAGCAATTCTCCTGCCTC 24152

Qy       1336 AGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTATT 1393
             ||||||| || || |||||| |||||||  |||  |  ||  || ||||||||||||||
Db      24153 AGCTTCCTGAGTACCTGGGACTACAGGCGTGTGCCACCACACCTGGCTAATTTTTGTATT 24212

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             |||| |||||||||  |||| |||||||| |||  || | || ||||||||||||||||
Db      24213 TTTAATAGAGACAGGATTTCACCATGTTGGCCAACCTGGTCTAGAACTCCTGACTTCAGG 24272
```

```
Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||| ||  ||||||||||||||||  ||||||||||||  |||||||| ||||||
Db      24273 TGATCCGCCTTCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCCA 24332

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTT 1541
              ||||| ||| ||   | ||   ||  ||
Db      24333 GCCTGCTTTTTCTTTTTCTTATTTATTT 24360


RESULT 41
ADH54059
ID   ADH54059 standard; DNA; 128034 BP.
XX
AC   ADH54059;
XX
DT   25-MAR-2004  (first entry)
XX
DE   Human IDE gene DNA sequence SeqID186.
XX
KW   human; neurodegenerative disease; urokinase plasminogen activator; uPA;
KW   gamma-synuclein; SNCG; insulin degrading enzyme; IDE;
KW   kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;
KW   tumour necrosis factor receptor SF6; TNFRSF6; Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(2456,G)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3279,C)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3407,T)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17095,G)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17242,T)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(33590,G)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(38903,T)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(42943,C)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(43391,G)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(45017,T)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(62498,C)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(68906,G)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(68973,T)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(69586,C)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(73772,G)
FT                   /*tag=  o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(74084,T)
FT                   /*tag=  p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(83024,T)
FT                   /*tag=  q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(83104,T)
FT                   /*tag=  r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(89301,G)
FT                   /*tag=  s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(105060,G)
FT                   /*tag=  t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(107395,A)
FT                   /*tag=  u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(108489,T)
FT                   /*tag=  v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(111914,G)
FT                   /*tag=  w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(112114,A)
FT                   /*tag=  x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(113142,G)
FT                   /*tag=  y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(113591,G)
FT                   /*tag=  z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(114683,G)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(116662,A)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(117803,T)
FT                   /*tag=  ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(124565,T)
FT                   /*tag=  ad
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   US2003224380-A1.
XX
PD   04-DEC-2003.
XX
PF   25-OCT-2002; 2002US-00282174.
XX
PR   25-OCT-2001; 2001US-0339525P.
PR   25-OCT-2001; 2001US-0348065P.
PR   02-NOV-2001; 2001US-0336983P.
PR   08-NOV-2001; 2001US-0336929P.
PR   08-NOV-2001; 2001US-0338010P.
PR   09-NOV-2001; 2001US-0338363P.
PR   04-DEC-2001; 2001US-0337052P.
PR   28-MAR-2002; 2002US-0368919P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Becker KD,  Velicelebi G,  Ellliott KJ,  Wang X,  Tanzi RE;
PI   Bertram L,  Saunders AJ,  Mullin KM,  Sampson AJ;
XX
DR   WPI; 2004-060538/06.
XX
PT   Determining a predisposition for or the occurrence of neurodegenerative
PT   disease, particularly Alzheimer's disease, comprises determining the
PT   presence of a polymorphism in the uPA, SNCG, IDE, KNSL1, LIPA or TNFRSF6
PT   gene.
XX
PS   Claim 22; SEQ ID NO 186; 205pp; English.
```

```
XX
CC   This invention relates to a novel method of determining a predisposition
CC   for or the occurrence of neurodegenerative disease comprising detecting
CC   in a target nucleic acid obtained from the subject the presence of an
CC   allelic variant of polymorphic regions of human genes selected from
CC   urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin
CC   degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid
CC   lipase (LIPA) and tumour necrosis factor receptor SF6 (TNFRSF6). The
CC   method is useful in determining the presence or predisposition to a
CC   neurodegenerative disease, particularly Alzheimer's disease. The present
CC   sequence is that of the human IDE gene which is related to the invention.
XX
SQ   Sequence 128034 BP; 34731 A; 25985 C; 26409 G; 40808 T; 0 U; 101 Other;

  Query Match             10.1%;  Score 201.6;  DB 12;  Length 128034;
  Best Local Similarity   82.0%;  Pred. No. 4.1e-40;
  Matches  269;  Conservative    0;  Mismatches   54;  Indels    5;  Gaps    3;

Qy       1218 TTTTTATTTTTATTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGT 1277
             | ||  ||||| |||||||||| |||||||||||| |||  |||||||||||| ||||
Db      24034 TATTCTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAGT 24093

Qy       1278 GGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TC 1335
             |||  | |||| |||||||||||||||||||||||| |||||||| |||||||||  ||
Db      24094 GGC-ATAGTCTTGGCTCACTGCAACCTCTGCCTCCCAAGTTCAAGCAATTCTCCTGCCTC 24152

Qy       1336 AGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTATT 1393
             ||||||| || || |||||| |||||||  |||  |  ||  || ||||||||||||||
Db      24153 AGCTTCCTGAGTACCTGGGACTACAGGCGTGTGCCACCACACCTGGCTAATTTTTGTATT 24212

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             |||| |||||||||  |||| |||||||| |||  || | || ||||||||||||||||
Db      24213 TTTAATAGAGACAGGATTTCACCATGTTGGCCAACCTGGTCTAGAACTCCTGACTTCAGG 24272

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||| ||  |||||||||||||||  |||||||||||  ||||||||| ||||||
Db      24273 TGATCCGCCTTCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCCA 24332

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             ||||| ||| ||   | ||   ||  ||
Db      24333 GCCTGCTTTTTCTTTTTCTTATTTATTT 24360


RESULT 42
ADH54060
ID   ADH54060 standard; DNA; 128034 BP.
XX
AC   ADH54060;
XX
DT   25-MAR-2004  (first entry)
XX
DE   Human IDE gene variant DNA sequence SeqID187.
XX
KW   human; neurodegenerative disease; urokinase plasminogen activator; uPA;
KW   gamma-synuclein; SNCG; insulin degrading enzyme; IDE;
KW   kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;
KW   tumour necrosis factor receptor SF6; TNFRSF6; Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
PN   US2003224380-A1.
XX
PD   04-DEC-2003.
XX
PF   25-OCT-2002; 2002US-00282174.
XX
PR   25-OCT-2001; 2001US-0339525P.
PR   25-OCT-2001; 2001US-0348065P.
PR   02-NOV-2001; 2001US-0336983P.
PR   08-NOV-2001; 2001US-0336929P.
PR   08-NOV-2001; 2001US-0338010P.
PR   09-NOV-2001; 2001US-0338363P.
PR   04-DEC-2001; 2001US-0337052P.
PR   28-MAR-2002; 2002US-0368919P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Becker KD,  Velicelebi G,  Ellliott KJ,  Wang X,  Tanzi RE;
PI   Bertram L,  Saunders AJ,  Mullin KM,  Sampson AJ;
XX
DR   WPI; 2004-060538/06.
XX
PT   Determining a predisposition for or the occurrence of neurodegenerative
PT   disease, particularly Alzheimer's disease, comprises determining the
PT   presence of a polymorphism in the uPA, SNCG, IDE, KNSL1, LIPA or TNFRSF6
PT   gene.
XX
PS   Claim 9; SEQ ID NO 187; 205pp; English.
XX
CC   This invention relates to a novel method of determining a predisposition
CC   for or the occurrence of neurodegenerative disease comprising detecting
CC   in a target nucleic acid obtained from the subject the presence of an
CC   allelic variant of polymorphic regions of human genes selected from
CC   urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin
CC   degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid
CC   lipase (LIPA) and tumour necrosis factor receptor SF6 (TNFRSF6). The
CC   method is useful in determining the presence or predisposition to a
CC   neurodegenerative disease, particularly Alzheimer's disease. The present
CC   sequence is that of the human IDE gene, with polymorphic sites
CC   represented by n, which is related to the invention.
XX
SQ   Sequence 128034 BP; 34726 A; 25977 C; 26400 G; 40799 T; 0 U; 132 Other;

  Query Match             10.1%;  Score 201.6;  DB 12;  Length 128034;
  Best Local Similarity   82.0%;  Pred. No. 4.1e-40;
  Matches  269;  Conservative    0;  Mismatches   54;  Indels    5;  Gaps    3;

Qy       1218 TTTTTATTTTTATTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGT 1277
             | ||  ||||| |||||||||| |||||||||||| |||  |||||||||||| ||||
Db      24034 TATTCTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAGT 24093

Qy       1278 GGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TC 1335
             |||  | |||| |||||||||||||||||||||||| |||||||| |||||||||  ||
Db      24094 GGC-ATAGTCTTGGCTCACTGCAACCTCTGCCTCCCAAGTTCAAGCAATTCTCCTGCCTC 24152

Qy       1336 AGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTATT 1393
             ||||||| || || |||||| |||||||  |||  |  ||  || ||||||||||||||
Db      24153 AGCTTCCTGAGTACCTGGGACTACAGGCGTGTGCCACCACACCTGGCTAATTTTTGTATT 24212

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             |||| |||||||||  |||| |||||||| |||  || | || ||||||||||||||||
Db      24213 TTTAATAGAGACAGGATTTCACCATGTTGGCCAACCTGGTCTAGAACTCCTGACTTCAGG 24272

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||| ||  |||||||||||||||  |||||||||||  ||||||||| ||||||
Db      24273 TGATCCGCCTTCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCCA 24332

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             ||||| ||| ||   | ||   ||  ||
Db      24333 GCCTGCTTTTTCTTTTTCTTATTTATTT 24360


RESULT 43
ADE43315/c
ID   ADE43315 standard; DNA; 202100 BP.
XX
AC   ADE43315;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human IDE/ KNSL1 genomic sequence, SEQ ID 484.
XX
KW   Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;
KW   Alzheimer's disease; neuroprotective; nootropic; gene therapy;
KW   Chromosome 10; gene; ds.
XX
```

OS Homo sapiens.
XX
PN WO2003054143-A2.
XX
PD 03-JUL-2003.
XX
PF 25-OCT-2002; 2002WO-US034679.
XX
PR 25-OCT-2001; 2001US-0339525P.
PR 08-NOV-2001; 2001US-0336929P.
PR 08-NOV-2001; 2001US-0338010P.
PR 09-NOV-2001; 2001US-0338363P.
PR 04-DEC-2001; 2001US-0337052P.
PR 28-MAR-2002; 2002US-0368919P.
XX
PA (NEUR-) NEUROGENETICS INC.
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Becker KD, Velicelebi G, Elliott KJ, Wang X, Tanzi RE, Bertram L;
PI Saunders AJ, Mullin KM, Sampson AJ, Blacker DL;
XX
DR WPI; 2003-559131/52.
XX
PT Determining a predisposition for or the occurrence of neurodegenerative
PT disease, e.g. Alzheimer's disease by detecting in a target nucleic acid
PT the presence or absence of an allelic variant of one or more polymorphic
PT regions.
XX
PS Claim 9; Page 769-823; 848pp; English.
XX
CC The present invention relates to a method (M1) for determining a
CC predisposition for or the occurrence of neurodegenerative disease in a
CC subject. The method comprises detecting in a target nucleic acid obtained
CC from the subject the presence or absence of an allelic variant of one or
CC more polymorphic regions of one or more genes selected from uPA
CC (Urokinase plasminogen activator), SNCG (gamma-synuclein), IDE (insulin-
CC degrading enzyme), KNSL1 (Kinesin-like protein 1), LIPA (lysosomal acid
CC lypase), and TNFRSF6 (Tumour Necrosis Factor Receptor-SF6), where the
CC presence of at least one of the allelic variant of one or more
CC polymorphic regions is indicative of a predisposition for or the
CC occurrence of neurodegenerative disease. The genes are all located on
CC chromosome 10. M1 is useful for determining a predisposition for or the
CC occurrence of, and for treating neurodegenerative disease, particularly
CC Alzheimer's disease.
XX
SQ Sequence 202100 BP; 60747 A; 41352 C; 41113 G; 58888 T; 0 U; 0 Other;

Query Match 10.1%; Score 201.6; DB 10; Length 202100;
Best Local Similarity 82.0%; Pred. No. 4.9e-40;
Matches 269; Conservative 0; Mismatches 54; Indels 5; Gaps 3;

```
Qy       1218 TTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGT 1277
             | ||  ||||| |||||||||| ||||||||||||| |||  |||||||||||| ||||
Db      99661 TATTCTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAGT 99602

Qy       1278 GGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TC 1335
              |||  | |||| |||||||||||||||||||||||| |||||||| |||||||||  ||
Db      99601 GGC-ATAGTCTTGGCTCACTGCAACCTCTGCCTCCCAAGTTCAAGCAATTCTCCTGCCTC 99543

Qy       1336 AGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTATT 1393
              ||||||| || || |||||| ||||||||  |||  |  ||  || |||||||||||||
Db      99542 AGCTTCCTGAGTACCTGGGACTACAGGCGTGTGCCACCACACCTGGCTAATTTTTGTATT 99483

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
              |||| |||||||||  |||| |||||||| |||  || | || ||||||||||||||||
Db      99482 TTTAATAGAGACAGGATTTCACCATGTTGGCCAACCTGGTCTAGAACTCCTGACTTCAGG 99423

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
              |||||| ||  |||||||||||||||  ||||||||||||  ||||||||| ||||||
Db      99422 TGATCCGCCTTCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCCA 99363

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTT 1541
              ||||| ||| ||   | ||   ||  ||
```

Db 99362 GCCTGCTTTTTCTTTTTCTTATTTATTT 99335

RESULT 44
ADH54357/c
ID ADH54357 standard; DNA; 202100 BP.
XX
AC ADH54357;
XX
DT 25-MAR-2004 (first entry)
XX
DE Human IDE/KNSL1 gene DNA sequence SeqID484.
XX
KW human; neurodegenerative disease; urokinase plasminogen activator; uPA;
KW gamma-synuclein; SNCG; insulin degrading enzyme; IDE;
KW kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;
KW tumour necrosis factor receptor SF6; TNFRSF6; Alzheimer's disease; ds.
XX
OS Homo sapiens.
XX
PN US2003224380-A1.
XX
PD 04-DEC-2003.
XX
PF 25-OCT-2002; 2002US-00282174.
XX
PR 25-OCT-2001; 2001US-0339525P.
PR 25-OCT-2001; 2001US-0348065P.
PR 02-NOV-2001; 2001US-0336983P.
PR 08-NOV-2001; 2001US-0336929P.
PR 08-NOV-2001; 2001US-0338010P.
PR 09-NOV-2001; 2001US-0338363P.
PR 04-DEC-2001; 2001US-0337052P.
PR 28-MAR-2002; 2002US-0368919P.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Becker KD, Velicelebi G, Ellliott KJ, Wang X, Tanzi RE;
PI Bertram L, Saunders AJ, Mullin KM, Sampson AJ;
XX
DR WPI; 2004-060538/06.
XX
PT Determining a predisposition for or the occurrence of neurodegenerative
PT disease, particularly Alzheimer's disease, comprises determining the
PT presence of a polymorphism in the uPA, SNCG, IDE, KNSL1, LIPA or TNFRSF6
PT gene.
XX
PS Claim 9; SEQ ID NO 484; 205pp; English.
XX
CC This invention relates to a novel method of determining a predisposition
CC for or the occurrence of neurodegenerative disease comprising detecting
CC in a target nucleic acid obtained from the subject the presence of an
CC allelic variant of polymorphic regions of human genes selected from
CC urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin
CC degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid
CC lipase (LIPA) and tumour necrosis factor receptor SF6 (TNFRSF6). The
CC method is useful in determining the presence or predisposition to a
CC neurodegenerative disease, particularly Alzheimer's disease. The present
CC sequence is that of a region of human DNA, containing the human IDE and
CC KNSL1 genes, which is related to the invention.
XX
SQ Sequence 202100 BP; 60747 A; 41352 C; 41113 G; 58888 T; 0 U; 0 Other;

Query Match 10.1%; Score 201.6; DB 12; Length 202100;
Best Local Similarity 82.0%; Pred. No. 4.9e-40;
Matches 269; Conservative 0; Mismatches 54; Indels 5; Gaps 3;

```
Qy       1218 TTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGT 1277
             | ||  ||||| |||||||||| ||||||||||||| |||  |||||||||||| ||||
Db      99661 TATTCTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAGT 99602

Qy       1278 GGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TC 1335
              |||  | |||| |||||||||||||||||||||||| |||||||| |||||||||  ||
```

```
Db      99601 GGC-ATAGTCTTGGCTCACTGCAACCTCTGCCTCCCAAGTTCAAGCAATTCTCCTGCCTC 99543

Qy       1336 AGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTATT 1393
              ||||||| || || |||||| |||||||| ||| | || || ||||||||||||||||
Db      99542 AGCTTCCTGAGTACCTGGGACTACAGGCGTGTGCCACCACACCTGGCTAATTTTTGTATT 99483

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
              |||| |||||||||  |||| |||||||| ||| || | || ||||||||||||||||||
Db      99482 TTTAATAGAGACAGGATTTCACCATGTTGGCCAACCTGGTCTAGAACTCCTGACTTCAGG 99423

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
              |||||| ||  |||||||||||||||  |||||||||||  ||||||||| ||||||
Db      99422 TGATCCGCCTTCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCCA 99363

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTT 1541
              ||||| ||| ||   | ||   ||  ||
Db      99362 GCCTGCTTTTTCTTTTTCTTATTTATTT 99335


RESULT 45
ABD32868/c
ID   ABD32868 standard; DNA; 238417 BP.
XX
AC   ABD32868;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD17-053.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 548; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 238417 BP; 57141 A; 58273 C; 60612 G; 62391 T; 0 U; 0 Other;

  Query Match            10.1%;  Score 201.6;  DB 13;  Length 238417;
  Best Local Similarity  82.0%;  Pred. No. 5.3e-40;
  Matches  269;  Conservative    0;  Mismatches   54;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
              ||||||| ||||| |||| ||||| || ||||||| |  ||||||||||||||||||| ||
Db     158308 TCTTTTTCTTTTTCTTTTCTTTGAGACGGAGTCTCGCCCTGTCACCCAGGCTGGAGTGCA 158249

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
                 ||| || | |||||||||||||| |||||||||||||||| ||||||||||||||
Db     158248 ATGG-TGCGATCTCGGCTCACTGTAACCTCTGCCTCCCAGGCTCAAGCGATTCTCCTGCC 158190

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAATTTTTGTA 1391
              ||||| |||||| |||||||||||||||||    ||  |  ||   | |||||| |||||
Db     158189 TCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCACACCCGGCTAATTCTTGTA 158130

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
              |||||||||||||||| ||||| |||||||| ||||||| | ||  || ||||||| |||
Db     158129 TTTTTAGTAGAGACAGGGTTTCACCATGTTGCCCAGGCTGGTCTCAAAATCCTGACCTCA 158070

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
              |||||||||||| ||||||||||||||||  |||||||||||||| |||||||||  ||
Db     158069 GGTGATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCACC 158010

Qy       1512 CAGCCTGTTTTCTCAGATCCTGTATTTG 1539
              |||||  |  ||   ||   | ||| ||
Db     158009 CAGCCAATCCTCCAGGAATTTCTATGTG 157982


Search completed: January 22, 2006, 04:09:12
Job time : 817.081 secs
```

This page intentionally left blank

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:39:03 ; Search time 245.225 Seconds
(without alignments)
14504.671 Million cell updates/sec

Title: US-09-728-552A-3_COPY_19000_21000
Perfect score: 2001
Sequence: 1 aatgtttggatttcaaaggt.....cagggacatgggtgcatagc 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	80595	3	US-09-078-294-3	Sequence 3, Appli
	2	1964.6	98.2	80246	3	US-09-078-294-4	Sequence 4, Appli
c	3	213.6	10.7	72048	3	US-09-949-002-600	Sequence 600, App
c	4	213.6	10.7	72048	3	US-09-949-002-684	Sequence 684, App
	5	208	10.4	88906	3	US-09-949-016-17468	Sequence 17468, A
	6	208	10.4	115963	3	US-09-949-016-12298	Sequence 12298, A
	7	208	10.4	144322	3	US-09-949-016-15316	Sequence 15316, A
c	8	206.8	10.3	98962	3	US-09-949-016-14133	Sequence 14133, A
c	9	206.8	10.3	102884	3	US-09-949-016-17100	Sequence 17100, A
	10	206.4	10.3	601	3	US-09-949-016-82418	Sequence 82418, A
	11	206.4	10.3	601	3	US-09-949-016-187588	Sequence 187588,
	12	206.4	10.3	92227	3	US-09-949-016-11929	Sequence 11929, A
	13	206.4	10.3	92232	3	US-09-949-016-15421	Sequence 15421, A
c	14	206.4	10.3	141560	3	US-09-949-016-16476	Sequence 16476, A
	15	206	10.3	601	3	US-09-949-016-130301	Sequence 130301,
c	16	206	10.3	97989	3	US-09-949-016-13208	Sequence 13208, A
c	17	205.6	10.3	23015	3	US-09-949-016-12807	Sequence 12807, A
c	18	205.6	10.3	23024	3	US-09-949-016-17529	Sequence 17529, A
	19	204.4	10.2	107827	3	US-09-949-016-15790	Sequence 15790, A
	20	204.4	10.2	264665	3	US-09-949-016-13747	Sequence 13747, A
c	21	203.8	10.2	31867	3	US-09-949-016-16874	Sequence 16874, A
c	22	203.8	10.2	31875	3	US-09-949-016-12864	Sequence 12864, A
	23	203.6	10.2	64024	3	US-09-949-016-17593	Sequence 17593, A
	24	203.4	10.2	88490	3	US-09-949-016-12758	Sequence 12758, A
	25	203.4	10.2	88736	3	US-09-949-016-14222	Sequence 14222, A
c	26	202.8	10.1	144596	3	US-09-949-016-11749	Sequence 11749, A
c	27	202.8	10.1	144596	3	US-09-949-016-13035	Sequence 13035, A
c	28	202.6	10.1	80269	3	US-09-949-016-15681	Sequence 15681, A
c	29	202.4	10.1	45418	3	US-09-949-016-15870	Sequence 15870, A
c	30	202.4	10.1	59828	3	US-09-949-016-16238	Sequence 16238, A
	31	202.4	10.1	129327	3	US-09-949-016-12257	Sequence 12257, A
	32	202.4	10.1	129327	3	US-09-949-016-15368	Sequence 15368, A
c	33	201.6	10.1	36532	3	US-09-949-016-16585	Sequence 16585, A
c	34	201.6	10.1	67620	3	US-09-949-016-16939	Sequence 16939, A
c	35	201.6	10.1	87734	3	US-09-949-016-17521	Sequence 17521, A
c	36	201.4	10.1	46347	3	US-09-949-016-13469	Sequence 13469, A
c	37	201.4	10.1	192506	3	US-09-949-016-15830	Sequence 15830, A
	38	201.2	10.1	601	3	US-09-949-016-169667	Sequence 169667,
	39	201.2	10.1	601	3	US-09-949-016-171021	Sequence 171021,
c	40	201.2	10.1	99748	3	US-09-949-016-11990	Sequence 11990, A
c	41	201.2	10.1	99749	3	US-09-949-016-16518	Sequence 16518, A
	42	201.2	10.1	162025	3	US-09-834-700-13	Sequence 13, Appl
	43	201.2	10.1	162025	3	US-09-834-700-14	Sequence 14, Appl
	44	201.2	10.1	162025	3	US-09-834-700-17	Sequence 17, Appl
	45	201.2	10.1	162025	3	US-09-834-700-18	Sequence 18, Appl
	46	201	10.0	601	3	US-09-949-016-203095	Sequence 203095,
c	47	201	10.0	1000	3	US-09-018-584A-32	Sequence 32, Appl
c	48	201	10.0	1000	3	US-09-784-423-32	Sequence 32, Appl
c	49	201	10.0	90923	3	US-09-949-002-623	Sequence 623, App
c	50	201	10.0	90925	3	US-09-949-002-789	Sequence 789, App
	51	200.6	10.0	25274	3	US-09-949-016-16682	Sequence 16682, A
c	52	200.6	10.0	96690	3	US-09-949-016-17103	Sequence 17103, A
	53	200.4	10.0	28063	3	US-09-949-016-16795	Sequence 16795, A
	54	200.4	10.0	39949	3	US-09-949-016-14210	Sequence 14210, A
c	55	200.4	10.0	49818	3	US-09-949-016-12458	Sequence 12458, A
c	56	200.4	10.0	49829	3	US-09-949-016-14011	Sequence 14011, A
	57	200.4	10.0	112132	3	US-09-741-150-3	Sequence 3, Appli
	58	200.4	10.0	112132	3	US-10-160-187-3	Sequence 3, Appli
c	59	200.2	10.0	7303	3	US-09-949-016-15998	Sequence 15998, A
	60	200.2	10.0	32278	3	US-09-949-016-14575	Sequence 14575, A
	61	200.2	10.0	36620	3	US-09-949-016-16150	Sequence 16150, A
	62	200.2	10.0	246444	3	US-09-949-016-13113	Sequence 13113, A
	63	200	10.0	601	3	US-09-949-016-77863	Sequence 77863, A
c	64	200	10.0	51539	3	US-09-949-002-677	Sequence 677, App
c	65	200	10.0	51540	3	US-09-949-002-819	Sequence 819, App
	66	200	10.0	59319	3	US-09-949-016-16115	Sequence 16115, A
	67	200	10.0	74804	3	US-09-949-016-15118	Sequence 15118, A
	68	200	10.0	131978	3	US-09-949-016-13117	Sequence 13117, A
	69	199.6	10.0	601	3	US-09-949-016-46996	Sequence 46996, A
	70	199.6	10.0	601	3	US-09-949-016-169666	Sequence 169666,
	71	199.6	10.0	421883	3	US-09-949-016-12557	Sequence 12557, A
c	72	199.4	10.0	22976	3	US-09-269-939A-19	Sequence 19, Appl
c	73	199.4	10.0	23187	3	US-09-499-522-1	Sequence 1, Appli
c	74	199.4	10.0	29629	3	US-09-729-995-3	Sequence 3, Appli
c	75	199.4	10.0	29629	3	US-10-135-689-3	Sequence 3, Appli
c	76	199.4	10.0	29629	3	US-10-690-617-3	Sequence 3, Appli
c	77	199.4	10.0	33885	3	US-09-949-016-16081	Sequence 16081, A
c	78	199.4	10.0	79858	3	US-09-949-016-16080	Sequence 16080, A
	79	199.4	10.0	90618	3	US-09-949-016-15964	Sequence 15964, A
c	80	199.4	10.0	325791	3	US-09-768-185A-1	Sequence 1, Appli
c	81	199.2	10.0	601	3	US-09-949-016-177638	Sequence 177638,
c	82	199.2	10.0	41684	3	US-09-536-059-1	Sequence 1, Appli
c	83	199.2	10.0	84132	3	US-09-949-016-16241	Sequence 16241, A
	84	199	9.9	601	3	US-09-949-016-166484	Sequence 166484,
c	85	199	9.9	36307	3	US-09-949-016-17372	Sequence 17372, A
c	86	199	9.9	91933	3	US-09-949-016-11855	Sequence 11855, A
c	87	199	9.9	91933	3	US-09-949-016-14628	Sequence 14628, A
	88	198.8	9.9	601	3	US-09-949-002-2865	Sequence 2865, Ap
	89	198.8	9.9	601	3	US-09-949-002-4871	Sequence 4871, Ap
c	90	198.8	9.9	18836	3	US-09-949-016-12654	Sequence 12654, A
c	91	198.8	9.9	18838	3	US-09-949-016-15633	Sequence 15633, A
c	92	198.8	9.9	45539	3	US-09-949-016-11920	Sequence 11920, A
c	93	198.8	9.9	45540	3	US-09-949-016-17199	Sequence 17199, A
	94	198.8	9.9	72992	3	US-09-949-016-17592	Sequence 17592, A
	95	198.8	9.9	90541	3	US-09-759-359A-3	Sequence 3, Appli
	96	198.8	9.9	90541	3	US-10-207-973-3	Sequence 3, Appli
	97	198.8	9.9	146039	3	US-09-949-016-12449	Sequence 12449, A

98	198.8	9.9	156651	3	US-09-949-016-17349	Sequence 17349, A
99	198.6	9.9	601	3	US-09-949-016-165314	Sequence 165314,
c 100	198.6	9.9	28884	3	US-09-949-016-16401	Sequence 16401, A

ALIGNMENTS

RESULT 1
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match 100.0%; Score 2001; DB 3; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19000 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 19059

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19060 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 19119

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19120 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 19179

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19180 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 19239

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19240 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 19299

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19300 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19359

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19360 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19419

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19420 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19479

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19480 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19539

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19540 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19599

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19600 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19659

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19660 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19719

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19720 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19779

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTTCAGGTGGGATGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19780 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTTCAGGTGGGATGG 19839

Qy        841 TAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGCT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19840 TAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGCT 19899

Qy        901 CTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19900 CTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCCC 19959

Qy        961 GTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19960 GTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTAG 20019

Qy       1021 AATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20020 AATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTAT 20079

Qy       1081 AGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20080 AGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTTT 20139

Qy       1141 AGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20140 AGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGA 20199

Qy       1201 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCACTTTGTCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20200 TTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCACTTTGTCAC 20259

Qy       1261 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20260 CCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCA 20319

Qy       1321 AGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20320 AGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCC 20379

Qy       1381 TAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20380 TAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAAC 20439

Qy       1441 TCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20440 TCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGA 20499

Qy       1501 GCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20500 GCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTA 20559

Qy       1561 TCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20560 TCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTT 20619

Qy       1621 TGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20620 TGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATGTT 20679

Qy       1681 ATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTAT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      20680 ATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTTAT 20739

Qy       1741 TGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTAAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20740 TGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTAAG 20799

Qy       1801 TGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTGAC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20800 TGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTGAC 20859

Qy       1861 TGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACATGC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20860 TGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACATGC 20919

Qy       1921 TGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACGCA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20920 TGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACGCA 20979

Qy       1981 ACAGGGACATGGGTGCATAGC 2001
              |||||||||||||||||||||
Db      20980 ACAGGGACATGGGTGCATAGC 21000


RESULT 2
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

  Query Match             98.2%;  Score 1964.6;  DB 3;  Length 80246;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 1992;  Conservative    0;  Mismatches    9;  Indels    2;  Gaps    2;

Qy          1 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18736 AATGTTTGGATTTCAAAGGTAACAAGAAGCTGGAAAACAACTACTGGCCGAGTCTGAGAG 18795

Qy         61 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18796 TTTCAGCGGAGACTGGTGCAGCCTTGTGTTTTTCCACTGACAGCTGAAAATGAGCCCAGC 18855

Qy        121 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18856 TTCAGTGAAGCTTGTTTCCTTCCCTCCTCAAGGTTACCCACAATTCTCAGTTCTCTCAGG 18915

Qy        181 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18916 AAAGCCAAAAAATGAATTTGAGGGTTTAGGATTGTGGTTCTTTTATCTATTACAGGATTG 18975

Qy        241 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18976 ATAATATGTTCCTCCACCAGATGTTCTGCTTGTAACAATACTCACTTCCTGACACTACTG 19035

Qy        301 CATATGCAGGAGTGTTACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 360
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      19036 CATATGCAGGAGTGTCACTACCAAGGTAAACACAGAATTGGCTGCCCAATTCCAAATCCC 19095

Qy        361 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      19096 TGAACTGAGTGAGAGAAATCAGAATTATAATAGGGGATTCAACAGAGCTGGCTACGGATG 19155

Qy        421 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19156 TGCCAGTGGTCAGATACTTTGCTCATCATACGCAGGTGCTGCTGCTCTAGCAACTGCTCA 19215

Qy        481 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19216 CTGCTTCATTTCCTGCCTTGGTCTTTAAATACTGCTTTTCTCAGCTCAATTGGCTTTCTT 19275

Qy        541 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19276 CCCTCTGGCAGTCACGTTTCTTTGGGTCAAACAGCAAATGATTCTTTAGAATCACCTGGT 19335

Qy        601 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19336 ACTCAAAGGAGCTACAAGACATTGGGCATCCACTTCCACTCTCTTGGAAAAACAATTTTA 19395

Qy        661 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19396 TGGAAGCCAAGGTTGCCATAGTGCCTCTTGAGGTTGTTTGCTCAGCCAAGGCCCAAGCTT 19455

Qy        721 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19456 TGTGCTTCAAACATGAAATTAGAGAGCTTCAGAACAAGATCCACATTTTCAATGGCCTCA 19515

Qy        781 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTT-TCAGGTGGGATG 839
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      19516 CCCAACTGGATAAAAGAACAATTGCCATATCTCAATGACCACCTTTTCTCAGGTGGGATG 19575

Qy        840 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19576 GTAGATGCTGGAATGGGTCACAGCATTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 19635

Qy        900 TCTGACTGGGGACCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAGAATATGAACCC 959
              ||||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db      19636 TCTGACTGGGGACCCTAAATATGCAAAAGTTAATAGGCTCTTCATTCAGAATATGAACCC 19695

Qy        960 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19696 CGTGTATGGATATAGCTAAAGGGTTGGCCTTTATGTTTCTATTCCTTCACAAACCTGGTA 19755

Qy       1020 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19756 GAATAGATATGCTTGTTTCCCTTTAAAAAATGTCAACAATTGCATTTATGATGCTGTGTA 19815

Qy       1080 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19816 TAGTAACTCACAGATCATGCTCCATGAAAATGCTTCAGAACCCAATATAAGGAGATTTTT 19875

Qy       1140 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19876 TAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTG 19935

Qy       1200 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCA 1259
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      19936 ATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGTGTCTCACTTTGTCA 19995

Qy       1260 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC 1319
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      19996 CCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTTTGCCTCCCAGGTTC 20055

Qy       1320 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC 1379
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      20056 AAGCGATTCTCCTGTCAGCTTCCCGAATAGCTGGGATTACAGGCGTATGCACCACCATGC 20115

Qy       1380 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAA 1439
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      20116 CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACTAGGCTTGCCTTGAA 20175

Qy       1440 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATG 1499
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      20176 CTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTTGGATTACAGGCATG 20235
```

```
Qy       1500 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA-TTTGTTTCTGAAGCCTTCATTTC 1558
             |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      20236 AGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTTGTTTCTGAAGCCTTCATTTC 20295

Qy       1559 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 1618
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20296 TATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATC 20355

Qy       1619 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 1678
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20356 TTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAGCTGATG 20415

Qy       1679 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 1738
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20416 TTATGTTTCTTTCAAATTATTCATTTCTCTATCTCAGAATTTATCTCATGCCTAATTGTT 20475

Qy       1739 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 1798
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20476 ATTGAATAGTCTTCACTTCTTGTCATCCAGTTTCTGGTCTCTTATTTCACTCTAAGTCTA 20535

Qy       1799 AGTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 1858
             | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20536 ATTGGCTATTAGAATAAAGAGCTTGTAACAGATTCTTTCTCCAATATGTCTTATCTTTTG 20595

Qy       1859 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 1918
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20596 ACTGCATGCCAGTGACAAACTGTTAACTGTTTTGATTCTTCATAACATTCCACAGAACAT 20655

Qy       1919 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 1978
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20656 GCTGACTCCTCTCTTCCTGAAAGCAATGCCCAAGCACAGCATTGTTAGATAGTATGTACG 20715

Qy       1979 CAACAGGGACATGGGTGCATAGC 2001
             |||||||||||||||||||||||
Db      20716 CAACAGGGACATGGGTGCATAGC 20738
```

RESULT 3
US-09-949-002-600/c
; Sequence 600, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 600
; LENGTH: 72048
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(72048)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-600

Query Match 10.7%; Score 213.6; DB 3; Length 72048;
Best Local Similarity 74.0%; Pred. No. 8.9e-47;
Matches 311; Conservative 0; Mismatches 104; Indels 5; Gaps 3;

```
Qy       1208 TTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCT 1267
             | ||   || | ||| | ||| |||||||||| ||  |||||| || ||||||||||||||
Db      64752 TACTTGTATTTCTTTTTCTTTTTTTTTTTTGAGACGAAGTCTCTCTCTGTCACCCAGGCT 64693

Qy       1268 GGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 1327
             ||||| |||||||    | |||| |||||||||||||||||||| ||||||||||||||||
Db      64692 GGAGTGCAGTGGC-ACGATCTCAGCTCACTGCAACCTCTGCCTCTCAGGTTCAAGCGATT 64634

Qy       1328 CTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAA 1383
             ||||||  | ||| |||||| |||||||||||||||| ||  ||  |  ||  || |||
Db      64633 CTCCTGCCTTAGCCTCCCGAGTAGCTGGGATTACAGGTGCCAGCCACCACACCTGGCTAC 64574

Qy       1384 TTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCC 1443
             |||||||||||||||||| ||||| |||||||||||||  |||||| | ||||||||||
Db      64573 TTTTTGTATTTTTAGTAGGGACAGGGTTTCGCCATGTTGGTCAGGCTGGTCTTGAACTCC 64514

Qy       1444 TGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCC 1503
             |||| |||||||||||||||||  ||||||||||  |||||| |||||||||| ||
Db      64513 TGACCTCAGGTGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATAACAGGCATGAACC 64454

Qy       1504 ACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
             ||||||||| |||   | |   |  || | | ||  ||          ||||   |||
Db      64453 ACCGTGCCCCGCCATGTCTTGTATTTCTTTTGTTAAATTTATTCCTAAGTATTTTATTCT 64394

Qy       1564 TCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGG 1623
             | ||| || | ||| || | | |    ||| |   | |||    ||   ||| | | |
Db      64393 TTTTAATCCTATTGTAAATGGGATTTTTAAAAATTATGTTTTTGGATTGTTTATTTCTAG 64334
```

RESULT 4
US-09-949-002-684/c
; Sequence 684, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 684
; LENGTH: 72048
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(72048)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-684

Query Match 10.7%; Score 213.6; DB 3; Length 72048;
Best Local Similarity 74.0%; Pred. No. 8.9e-47;
Matches 311; Conservative 0; Mismatches 104; Indels 5; Gaps 3;

```
Qy       1208 TTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCT 1267
             | ||   || | ||| | ||| |||||||||| ||  |||||| || ||||||||||||||
Db      64752 TACTTGTATTTCTTTTTCTTTTTTTTTTTTGAGACGAAGTCTCTCTCTGTCACCCAGGCT 64693

Qy       1268 GGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 1327
             ||||| |||||||    | |||| |||||||||||||||||||| ||||||||||||||||
Db      64692 GGAGTGCAGTGGC-ACGATCTCAGCTCACTGCAACCTCTGCCTCTCAGGTTCAAGCGATT 64634

Qy       1328 CTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA--TGCCTAA 1383
             ||||||  | ||| |||||| |||||||||||||||| ||  ||  |  ||  || |||
Db      64633 CTCCTGCCTTAGCCTCCCGAGTAGCTGGGATTACAGGTGCCAGCCACCACACCTGGCTAC 64574

Qy       1384 TTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCC 1443
             |||||||||||||||||| ||||| |||||||||||||  |||||| | ||||||||||
Db      64573 TTTTTGTATTTTTAGTAGGGACAGGGTTTCGCCATGTTGGTCAGGCTGGTCTTGAACTCC 64514

Qy       1444 TGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCC 1503
             |||| |||||||||||||||||  ||||||||||  |||||| |||||||||| ||
Db      64513 TGACCTCAGGTGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATAACAGGCATGAACC 64454
```

```
Qy       1504 ACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
             ||||||||| |||   | |   |  || | | ||   ||            ||||   |||
Db      64453 ACCGTGCCCCGCCATGTCTTGTATTTCTTTTGTTAAATTTATTCCTAAGTATTTTATTCT 64394

Qy       1564 TCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCAAATATCTTTGG 1623
             | ||| || | ||| || | | |    ||| |   | |||    ||    ||| | | |
Db      64393 TTTTAATCCTATTGTAAATGGGATTTTTAAAAATTATGTTTTTGGATTGTTTATTTCTAG 64334


RESULT 5
US-09-949-016-17468
; Sequence 17468, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17468
;   LENGTH: 88906
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(88906)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17468

  Query Match            10.4%;  Score 208;  DB 3;  Length 88906;
  Best Local Similarity   78.0%;  Pred. No. 3.2e-45;
  Matches  276;  Conservative    0;  Mismatches   75;  Indels    3;  Gaps    2;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             | ||||| ||||| ||||||| || |  ||||||  || ||||||||||||||||||| ||
Db      36054 TTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCA 36113

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||| |||  |||| ||||| |||||||||||||||||||||||| ||||||||||
Db      36114 GTGG-TGTAATCTCAGCTCAGCGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGCC 36172

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGCCTAATTTTTGTATT 1393
             ||||| |||| | ||||||||| ||||||| | |||  |  || ||       ||||||||
Db      36173 TCAGCCTCCCAAGTAGCTGGGACTACAGGCACGTGCCACCACACGCGGCTAATTTGTATT 36232

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             ||||||||||||| |||||||||||| ||||||| | |||||||||||| |||||
Db      36233 TTTAGTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGG 36292

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||||||||||| |||||||||||  |||||||||||||||||||||| |  |||
Db      36293 TGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCTCCCG 36352

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTT 1567
             ||| |  | |     |  |     |  || |    || |  ||| | |||| ||
Db      36353 GCCCGGAATTTTCTTTATTTCTATACATTTTACTTCCGTTGTTTTTTTCTTTTT 36406


RESULT 6
US-09-949-016-12298
; Sequence 12298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12298
;   LENGTH: 115963
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(115963)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12298

  Query Match            10.4%;  Score 208;  DB 3;  Length 115963;
  Best Local Similarity   79.2%;  Pred. No. 3.6e-45;
  Matches  285;  Conservative    0;  Mismatches   70;  Indels    5;  Gaps    3;

Qy       1187 AGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAG 1246
             |||||    |  |||||  ||  |    || ||   || ||| |||||||||||||| ||
Db       2424 AGAGACACTTCACATTATATTAACATTAATGTTAACATCTTTTTTTTTTTTGAAACGGAA 2483

Qy       1247 TCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCT 1306
             ||||||  |||  |||||||||||||||||||| |||| |||||||||||||| ||| |
Db       2484 TCTCACACTGTTGCCCAGGCTGGAGTACAGTGG-TGTGATCTCGGCTCACTGCAGCCTTT 2542

Qy       1307 GCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCG 1364
             |||||||||||||||| |||||||||  ||||| ||| ||||||||||||||||||||
Db       2543 GCCTCCCAGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAATAGCTGGGATTACAGGCA 2602

Qy       1365 CATGC--ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTG 1422
             |||||   |   | || |||||||||||||||||||||||||||| |||| | |||||||
Db       2603 CATGCCACCACGCCTGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTTGTCATGTTG 2662

Qy       1423 ACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCA 1482
              |||||||  | ||  | |||||||  |||||||||||| ||||||||||||||||||||
Db       2663 GCCAGGCTGGTCTCAAGCTCCTGATGTCAGGTGATCCGCCCACCTCAGCCTCCCAAAGTG 2722

Qy       1483 CTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTT 1542
             |||||||||||||||||||||| |||||  |||  |    ||  ||  || |||   ||
Db       2723 CTGGGATTACAGGCATGAGCCACTGTGCCTGGCCATTAACATCTAATTATGCATTGAGTT 2782


RESULT 7
US-09-949-016-15316
; Sequence 15316, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15316
```

```
;   LENGTH: 144322
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(144322)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15316

  Query Match             10.4%;  Score 208;  DB 3;  Length 144322;
  Best Local Similarity   79.2%;  Pred. No. 4e-45;
  Matches  285;  Conservative    0;  Mismatches   70;  Indels    5;  Gaps    3;

Qy       1187 AGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAG 1246
              |||||     |   |||||  ||  |    || ||    || ||| ||||||||||||| ||
Db      78783 AGAGACACTTCACATTATATTAACATTAATGTTAACATCTTTTTTTTTTTTGAAACGGAA 78842

Qy       1247 TCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCT 1306
              ||||||  |||  |||||||||||||||||||| |||| |||||||||||||| ||| |
Db      78843 TCTCACACTGTTGCCCAGGCTGGAGTACAGTGG-TGTGATCTCGGCTCACTGCAGCCTTT 78901

Qy       1307 GCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCG 1364
              |||||||||||||||| |||||||||  |||||  ||| |||||||||||||||||||
Db      78902 GCCTCCCAGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAATAGCTGGGATTACAGGCA 78961

Qy       1365 CATGC--ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTG 1422
              |||||    |    | || ||||||||||||||||||||||||||||| |||| | |||||||
Db      78962 CATGCCACCACGCCTGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTTGTCATGTTG 79021

Qy       1423 ACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCA 1482
              |||||||  | ||   | |||||||  |||||||||| ||||||||||||||||||||
Db      79022 GCCAGGCTGGTCTCAAGCTCCTGATGTCAGGTGATCCGCCCACCTCAGCCTCCCAAAGTG 79081

Qy       1483 CTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTT 1542
              |||||||||||||||||||||||||| |||||  |||  |    ||  ||  || |||   ||
Db      79082 CTGGGATTACAGGCATGAGCCACTGTGCCTGGCCATTAACATCTAATTATGCATTGAGTT 79141


RESULT 8
US-09-949-016-14133/c
; Sequence 14133, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14133
;   LENGTH: 98962
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(98962)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14133

  Query Match             10.3%;  Score 206.8;  DB 3;  Length 98962;
  Best Local Similarity   74.6%;  Pred. No. 7.1e-45;
  Matches  300;  Conservative    0;  Mismatches   97;  Indels    5;  Gaps    3;

Qy       1166 ATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATT 1225
              | ||| ||||  |  || |  ||| ||    |  | ||     |   ||    || | ||

Db      82606 AATTCTGTGTGTATCTTCTAGAGATAATCTATCTAATAAAGAATGGAAGCATATTCTTTT 82547

Qy       1226 TTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGG 1285
              |||||||||||||| |||||||||  || |||||||||||||||| | |||||||
Db      82546 TTTATTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAATGCAGTGGC-ACAA 82488

Qy       1286 TCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGTC--AGCTTCCC 1343
              |||| |||||||||||||||||||||||| ||||||  ||||||||||||   ||| || |
Db      82487 TCTCAGCTCACTGCAACCTCTGCCTCCCAGTTTCAAGAGATTCTCCTGTCTTAGCCTCTC 82428

Qy       1344 GAATAGCTGGGATTACAGGCG--CATGCACCACCATGCCTAATTTTTGTATTTTTAGTAG 1401
              || ||||||| ||||||||||  ||     |   | || ||||||||||||||||||| ||
Db      82427 GAGTAGCTGGAATTACAGGCGTCCACCACCACGCCTGGCTAATTTTTGTATTTTTAGCAG 82368

Qy       1402 AGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCAC 1461
              ||||  ||||| |||||||| ||||||| | ||||||||||||||| ||||||||||||
Db      82367 TTACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAC 82308

Qy       1462 CCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
              |  |||  |||||||||| |||||||||||||| ||||||| ||||||| || |||
Db      82307 CTGCCTTGGCCTCCCAAAGTACTGGGATTACAGGCGTGAGCCACTGTGCCCAGTCTTTTT 82248

Qy       1522 TCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
              | |    |  | | ||| |||    |   | |||  || | |
Db      82247 TTTTTTTTTTTTTTTTTTTTTGAGATGGTACATCCCTCTGT 82206


RESULT 9
US-09-949-016-17100/c
; Sequence 17100, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17100
;   LENGTH: 102884
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(102884)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17100

  Query Match             10.3%;  Score 206.8;  DB 3;  Length 102884;
  Best Local Similarity   74.6%;  Pred. No. 7.2e-45;
  Matches  300;  Conservative    0;  Mismatches   97;  Indels    5;  Gaps    3;

Qy       1166 ATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATT 1225
              | ||| ||||  |  || |  ||| ||    |  | ||     |   ||    || | ||

Db      82606 AATTCTGTGTGTATCTTCTAGAGATAATCTATCTAATAAAGAATGGAAGCATATTCTTTT 82547

Qy       1226 TTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGG 1285
              |||||||||||||| |||||||||  || |||||||||||||||| | |||||||
Db      82546 TTTATTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAATGCAGTGGC-ACAA 82488

Qy       1286 TCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGTC--AGCTTCCC 1343
              |||| |||||||||||||||||||||||| ||||||  ||||||||||||   ||| || |
Db      82487 TCTCAGCTCACTGCAACCTCTGCCTCCCAGTTTCAAGAGATTCTCCTGTCTTAGCCTCTC 82428

Qy       1344 GAATAGCTGGGATTACAGGCG--CATGCACCACCATGCCTAATTTTTGTATTTTTAGTAG 1401
```

```
              || ||||||| |||||||||  ||    |   | || ||||||||||||||||||| ||
Db      82427 GAGTAGCTGGAATTACAGGCGTCCACCACCACGCCTGGCTAATTTTTGTATTTTTAGCAG 82368

Qy       1402 AGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCAC 1461
              |||| ||||| |||||||| ||||||| | |||||||||||||| |||||||||||||
Db      82367 TTACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAC 82308

Qy       1462 CCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
              |  |||  |||||||||||| |||||||||||||||| |||||||| |||||||| || |||
Db      82307 CTGCCTTGGCCTCCCAAAGTACTGGGATTACAGGCGTGAGCCACTGTGCCCAGTCTTTTT 82248

Qy       1522 TCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
              | |    |  | | ||| |||    |   | |||  || | |
Db      82247 TTTTTTTTTTTTTTTTTTTTTTGAGATGGTACATCCCTCTGT 82206


RESULT 10
US-09-949-016-82418
; Sequence 82418, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 82418
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-82418

  Query Match             10.3%;  Score 206.4;  DB 3;  Length 601;
  Best Local Similarity   74.4%;  Pred. No. 7.7e-46;
  Matches  299;  Conservative    1;  Mismatches   97;  Indels    5;  Gaps    3;

Qy       1166 ATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATT 1225
              | ||| ||||  |  || |  ||| ||    |  | ||      |   ||    || | ||
Db        181 AATTCTGTGTGTATCTTCTAGAGATAATCTATCTAATAAAGAATGGAAGCATATTCTTTT 240

Qy       1226 TTTATTTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGG 1285
              |||||||||||||| |||||||||  || |||||||||||||||| | |||||||
Db        241 TTTATTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAATGCAGTGGC-ACAA 299

Qy       1286 TCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGTC--AGCTTCCC 1343
              |:|| |||||||||||||||||||||||||| |||||| ||||||||||||  ||| || |
Db        300 TYTCAGCTCACTGCAACCTCTGCCTCCCAGTTTCAAGAGATTCTCCTGTCTTAGCCTCTC 359

Qy       1344 GAATAGCTGGGATTACAGGCG--CATGCACCACCATGCCTAATTTTTGTATTTTTAGTAG 1401
              || ||||||| |||||||||  ||    |   | || ||||||||||||||||||| ||
Db        360 GAGTAGCTGGAATTACAGGCGTCCACCACCACGCCTGGCTAATTTTTGTATTTTTAGCAG 419

Qy       1402 AGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCAC 1461
              |||| ||||| |||||||| ||||||| | |||||||||||||| |||||||||||||
Db        420 TTACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAC 479

Qy       1462 CCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
              |  |||  |||||||||||| |||||||||||||||| |||||||| |||||||| || |||
Db        480 CTGCCTTGGCCTCCCAAAGTACTGGGATTACAGGCGTGAGCCACTGTGCCCAGTCTTTTT 539

Qy       1522 TCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
              | |    |  | | ||| |||    |   | |||  || | |
Db        540 TTTTTTTTTTTTTTTTTTTTTTGAGATGGTACATCCCTCTGT 581


RESULT 11
US-09-949-016-187588
; Sequence 187588, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 187588
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-187588

  Query Match             10.3%;  Score 206.4;  DB 3;  Length 601;
  Best Local Similarity   74.4%;  Pred. No. 7.7e-46;
  Matches  299;  Conservative    1;  Mismatches   97;  Indels    5;  Gaps    3;

Qy       1166 ATTTCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATT 1225
              | ||| ||||  |  || |  ||| ||    |  | ||      |   ||    || | ||
Db        181 AATTCTGTGTGTATCTTCTAGAGATAATCTATCTAATAAAGAATGGAAGCATATTCTTTT 240

Qy       1226 TTTATTTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGG 1285
              |||||||||||||| |||||||||  || |||||||||||||||| | |||||||
Db        241 TTTATTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAATGCAGTGGC-ACAA 299

Qy       1286 TCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGTC--AGCTTCCC 1343
              |:|| |||||||||||||||||||||||||| |||||| ||||||||||||  ||| || |
Db        300 TYTCAGCTCACTGCAACCTCTGCCTCCCAGTTTCAAGAGATTCTCCTGTCTTAGCCTCTC 359

Qy       1344 GAATAGCTGGGATTACAGGCG--CATGCACCACCATGCCTAATTTTTGTATTTTTAGTAG 1401
              || ||||||| |||||||||  ||    |   | || ||||||||||||||||||| ||
Db        360 GAGTAGCTGGAATTACAGGCGTCCACCACCACGCCTGGCTAATTTTTGTATTTTTAGCAG 419

Qy       1402 AGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCAC 1461
              |||| ||||| |||||||| ||||||| | |||||||||||||| |||||||||||||
Db        420 TTACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAC 479

Qy       1462 CCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTT 1521
              |  |||  |||||||||||| |||||||||||||||| |||||||| |||||||| || |||
Db        480 CTGCCTTGGCCTCCCAAAGTACTGGGATTACAGGCGTGAGCCACTGTGCCCAGTCTTTTT 539

Qy       1522 TCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCT 1563
              | |    |  | | ||| |||    |   | |||  || | |
Db        540 TTTTTTTTTTTTTTTTTTTTTTGAGATGGTACATCCCTCTGT 581


RESULT 12
US-09-949-016-11929
; Sequence 11929, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
```

```
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11929
;   LENGTH: 92227
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(92227)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11929

  Query Match             10.3%;  Score 206.4;  DB 3;  Length 92227;
  Best Local Similarity   74.0%;  Pred. No. 8.8e-45;
  Matches  302;  Conservative    0;  Mismatches  101;  Indels    5;  Gaps    3;

Qy       1198 TGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGT 1257
             | |||    | ||    | |    |||| ||||| ||||||| || ||| |||||||||| |||
Db      15018 TTATTTGATCTTTGGATAATCTTTTTTTTTTTTTTTTTTTGGAGACAAAGTCTCACTCTGT 15077

Qy       1258 CACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGT 1317
             |||||||| |||||  |||||| |   ||||||||||||||| ||||||||||||||||
Db      15078 CACCCAGGCCGGAGTGAAGTGGC-GCAATCTCGGCTCACTGCAGCCTCTGCCTCCCAGGT 15136

Qy       1318 TCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGC--GCATGCACCA 1373
             || |||||||||||||  |||||  ||||   |||||||  |||||  ||  |||
Db      15137 TCCAGCGATTCTCCTGCCTCAGCCTCCCAGTTAGCTGGTGTTACAAGCATGCACCACTAC 15196

Qy       1374 CCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGC 1433
               | || |||||||||||||||||||||||||| | ||||| ||||||||||||||||| |
Db      15197 GCCTGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGACCAGGCTGGT 15256

Qy       1434 CTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACA 1493
             || |||||||||||  ||||||||||  | | ||||  ||||||||||  || ||||||||
Db      15257 CTCGAACTCCTGACCTCAGGTGATCTGCTCGCCTCGGCCTCCCAAAGTGCTAGGATTACA 15316

Qy       1494 GGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTC 1553
             |||||||||||| | |||| ||||    ||| |||||  |   ||  | |   ||   ||
Db      15317 GGCATGAGCCACTGCGCCCGGCCTAAAATCTTAGATCTTTAGTTAATCTTATAACAATTT 15376

Qy       1554 ATTTCTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTT 1601
             ||    || |||||  |  |||||    || |      ||| |||| ||
Db      15377 ATAAAAATATTCTTCCTAGTTTTGCTCCTACTCTGAGTAAATAAGCTT 15424


RESULT 13
US-09-949-016-15421
; Sequence 15421, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15421
;   LENGTH: 92232
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(92232)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15421

  Query Match             10.3%;  Score 206.4;  DB 3;  Length 92232;
  Best Local Similarity   74.0%;  Pred. No. 8.8e-45;
  Matches  302;  Conservative    0;  Mismatches  101;  Indels    5;  Gaps    3;

Qy       1198 TGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGT 1257
             | |||    | ||    | |    |||| ||||| ||||||| || ||| |||||||||| |||
Db      15018 TTATTTGATCTTTGGATAATCTTTTTTTTTTTTTTTTTTTGGAGACAAAGTCTCACTCTGT 15077

Qy       1258 CACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGT 1317
             |||||||| |||||  |||||| |   ||||||||||||||| ||||||||||||||||
Db      15078 CACCCAGGCCGGAGTGAAGTGGC-GCAATCTCGGCTCACTGCAGCCTCTGCCTCCCAGGT 15136

Qy       1318 TCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGC--GCATGCACCA 1373
             || |||||||||||||  |||||  ||||   |||||||  |||||  ||  |||
Db      15137 TCCAGCGATTCTCCTGCCTCAGCCTCCCAGTTAGCTGGTGTTACAAGCATGCACCACTAC 15196

Qy       1374 CCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGC 1433
               | || |||||||||||||||||||||||||| | ||||| ||||||||||||||||| |
Db      15197 GCCTGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGACCAGGCTGGT 15256

Qy       1434 CTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACA 1493
             || |||||||||||  ||||||||||  | | ||||  ||||||||||  || ||||||||
Db      15257 CTCGAACTCCTGACCTCAGGTGATCTGCTCGCCTCGGCCTCCCAAAGTGCTAGGATTACA 15316

Qy       1494 GGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTC 1553
             |||||||||||| | |||| ||||    ||| |||||  |   ||  | |   ||   ||
Db      15317 GGCATGAGCCACTGCGCCCGGCCTAAAATCTTAGATCTTTAGTTAATCTTATAACAATTT 15376

Qy       1554 ATTTCTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTT 1601
             ||    || |||||  |  |||||    || |      ||| |||| ||
Db      15377 ATAAAAATATTCTTCCTAGTTTTGCTCCTACTCTGAGTAAATAAGCTT 15424


RESULT 14
US-09-949-016-16476/c
; Sequence 16476, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16476
;   LENGTH: 141560
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(141560)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16476

  Query Match             10.3%;  Score 206.4;  DB 3;  Length 141560;
  Best Local Similarity   76.3%;  Pred. No. 1.1e-44;
  Matches  293;  Conservative    0;  Mismatches   86;  Indels    5;  Gaps    3;

Qy       1195 ATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTT 1254
             | | || |   |||| ||   || |||| ||||| |||||||||| || ||||||||||
Db      66406 AATGGAATCGTTTTTTTCTTTTCCTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCACTC 66347
```

```
Qy       1255 TGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCA 1314
             ||||||||||||||||||| ||||||| |    |||| ||| ||||||||| |||||||||
Db      66346 TGTCACCCAGGCTGGAGTGCAGTGGC-GCTATCTCAGCTTACTGCAACCCCTGCCTCCCG 66288

Qy       1315 GGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAG--GCGCATGCA 1370
             ||||||| ||||||||||  ||||| |||| | ||||||||||||||  | |||
Db      66287 GGTTCAAATGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGTGCACCAC 66228

Qy       1371 CCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
             |   | || ||||||||||||||||||||||||||| | ||| |||||||| |||||||
Db      66227 CATGCCTGGCTAATTTTTGTATTTTTAGTAGAGACAGGGGTTCACCATGTTGGCCAGGCT 66168

Qy       1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490
              | |||||||||||||| ||| |||||||||| |||  | |||  ||||  ||||||||
Db      66167 GGTCTTGAACTCCTGACCTCAAGTGATCCACCCCCCTTGGGCTCTTAAAGTGCTGGGATT 66108

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCC 1550
             |||||| |||||||||||||||||| | ||||  | ||  | | | |  |||  ||
Db      66107 ACAGGCGTGAGCCACCGTGCCCAGCCGGATTTCAAAAATAATATCTGTCATTCCAAAATA 66048

Qy       1551 TTCATTTCTATCTTCTTATTCATT 1574
              |  ||       |  || ||| ||
Db      66047 GTACTTAGGTGATCATTTTTCTTT 66024


RESULT 15
US-09-949-016-130301
; Sequence 130301, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 130301
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-130301

  Query Match             10.3%;  Score 206;  DB 3;  Length 601;
  Best Local Similarity   73.8%;  Pred. No. 9.9e-46;
  Matches  301;  Conservative    1;  Mismatches  101;  Indels    5;  Gaps    3;

Qy       1198 TGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGT 1257
             | |||   | ||  | |   |||| ||||| ||||||| || ||| |||||||| |||
Db         59 TTATTTGATCTTTGGATAATCTTTTTTTTTTTTTTTTTTGGAGACAAAGTCTCACTCTGT 118

Qy       1258 CACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGT 1317
             ||||||||| |||||  |||||| |   |||||||||||||| ||||||||||||||||
Db        119 CACCCAGGCCGGAGTGAAGTGGC-GCAATCTCGGCTCACTGCAGCCTCTGCCTCCCAGGT 177

Qy       1318 TCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGC--GCATGCACCA 1373
             || |||||||||||||  ||||| ||||   |||||||  ||||| ||  |||
Db        178 TCCAGCGATTCTCCTGCCTCAGCCTCCCAGTTAGCTGGTGTTACAAGCATGCACCACTAC 237

Qy       1374 CCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGC 1433
              | || |||||||||||||||||||||||||| | ||||| ||||||||||||||| |
Db        238 GCCTGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGACCAGGCTGGT 297

Qy       1434 CTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACA 1493
             || :|||||||||| |||||||||||  | | |||| ||||||||||  || ||||||||
Db        298 CTCKAACTCCTGACCTCAGGTGATCTGCTCGCCTCGGCCTCCCAAAGTGCTAGGATTACA 357

Qy       1494 GGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTC 1553
             |||||||||||| | |||| ||||    ||| ||||| |   ||  | |  ||  ||
Db        358 GGCATGAGCCACTGCGCCCGGCCTAAAATCTTAGATCTTTAGTTAATCTTATAACAATTT 417

Qy       1554 ATTTCTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTT 1601
              ||    || |||||  |  |||||    || |     ||| |||| ||
Db        418 ATAAAAATATTCTTCCTAGTTTTGCTCCTACTCTGAGTAAATAAGCTT 465


RESULT 16
US-09-949-016-13208/c
; Sequence 13208, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13208
;   LENGTH: 97989
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(97989)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13208

  Query Match             10.3%;  Score 206;  DB 3;  Length 97989;
  Best Local Similarity   80.0%;  Pred. No. 1.2e-44;
  Matches  280;  Conservative    0;  Mismatches   65;  Indels    5;  Gaps    3;

Qy       1181 TTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAA 1240
             || |||      | |  || | ||  ||||      || | ||||| |||  |||||||||
Db      73215 TTATTCTATTTATTTATTAAATAATTTTTAAATTATTTATTTATATTTTATTTTTTTGAG 73156

Qy       1241 ACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCA 1300
             |||||||||| || |||||||||| ||||||||||||||||      |||| ||||||||||
Db      73155 ACAGAGTCTCGCTGTGTCACCCAGACTGGAGTACAGTGGC-ACAATCTCAGCTCACTGCA 73097

Qy       1301 ACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTA 1358
             |||||||||||||||||||||||||||||||| |||  ||||| |||| | ||||||||||||
Db      73096 ACCTCTGCCTCCCAGGTTCAAGCGATTCTTCTGCCTCAGCCTCCCAAGTAGCTGGGATTA 73037

Qy       1359 CAGGCGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCC 1416
             |||| ||  ||      ||  |  || |||||| |||||||||||||||||| ||||||||
Db      73036 CAGGTGCCCGCCATCACACCCAGCTCATTTTTATATTTTTAGTAGAGACAGGGTTTCGCC 72977

Qy       1417 ATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCC 1476
             ||||||  |||||| | || |||||||||| ||||||||| | | ||||||||||||
Db      72976 ATGTTGGTCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCTCTCGCCTCAGCCTCCC 72917

Qy       1477 AAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCA 1526
             |||| ||||||||||||||  |||||||||||||||||||| || | | |
Db      72916 AAAGTACTGGGATTACAGGTGTGAGCCACCGTGCCCAGCCTCTTGTTTTA 72867


RESULT 17
US-09-949-016-12807/c
; Sequence 12807, Application US/09949016
; Patent No. 6812339
```

```
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12807
;   LENGTH: 23015
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12807

  Query Match             10.3%;  Score 205.6;  DB 3;  Length 23015;
  Best Local Similarity   78.8%;  Pred. No. 7.4e-45;
  Matches  271;  Conservative    0;  Mismatches   69;  Indels    4;  Gaps    2;

Qy       1202 TATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACC 1261
             || |   | |||      |||| |||||  ||||| ||| ||||||||||||||||||||
Db       3249 TAAGGGATATCATCAGATTTTTTTTTTCCTTTTTCTGAGACAGAGTCTCACTTTGTCACC 3190

Qy       1262 CAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAA 1321
             ||||||||||||||||||   | |||||||||||||||||||| ||||| |||||||
Db       3189 CAGGCTGGAGTACAGTGGCGCAGATCTCGGCTCACTGCAACCTCTGGCTCCCGGGTTCAA 3130

Qy       1322 GCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCAT 1377
             | ||||||||||  ||||| ||| || |||||||||||||||||||| |||   |   |
Db       3129 GTGATTCTCCTGCCTCAGCCTCCGGAGTAGCTGGGATTACAGGCGCCTGCCACCATGCCC 3070

Qy       1378 GCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTG 1437
             | ||| |||||||||||||||||||| | ||||| |||||||  |||||||   |
Db       3069 GGCTACTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTAGCCAGGCTGTTCCCA 3010

Qy       1438 AACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCA 1497
             |||||||||| |||||  ||||  || ||||||||||||||  ||||||||||||||
Db       3009 AACTCCTGACCTCAGGCAATCCGGCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCG 2950

Qy       1498 TGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             |||||||||| |||| ||| |  ||  |   |  | | ||| ||
Db       2949 TGAGCCACCGAGCCCGGCCAGACTTGACTTTTTTTTTTTTTTTT 2906


RESULT 18
US-09-949-016-17529/c
; Sequence 17529, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17529
;   LENGTH: 23024
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17529

  Query Match             10.3%;  Score 205.6;  DB 3;  Length 23024;
  Best Local Similarity   78.8%;  Pred. No. 7.4e-45;
  Matches  271;  Conservative    0;  Mismatches   69;  Indels    4;  Gaps    2;

Qy       1202 TATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACC 1261
             || |   | |||      |||| |||||  ||||| ||| ||||||||||||||||||||
Db       3249 TAAGGGATATCATCAGATTTTTTTTTTCCTTTTTCTGAGACAGAGTCTCACTTTGTCACC 3190

Qy       1262 CAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAA 1321
             ||||||||||||||||||   | |||||||||||||||||||| ||||| |||||||
Db       3189 CAGGCTGGAGTACAGTGGCGCAGATCTCGGCTCACTGCAACCTCTGGCTCCCGGGTTCAA 3130

Qy       1322 GCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCAT 1377
             | ||||||||||  ||||| ||| || |||||||||||||||||||| |||   |   |
Db       3129 GTGATTCTCCTGCCTCAGCCTCCGGAGTAGCTGGGATTACAGGCGCCTGCCACCATGCCC 3070

Qy       1378 GCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTG 1437
             | ||| |||||||||||||||||||| | ||||| |||||||  |||||||   |
Db       3069 GGCTACTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTAGCCAGGCTGTTCCCA 3010

Qy       1438 AACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCA 1497
             |||||||||| |||||  ||||  || ||||||||||||||  ||||||||||||||
Db       3009 AACTCCTGACCTCAGGCAATCCGGCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCG 2950

Qy       1498 TGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTT 1541
             |||||||||| |||| ||| |  ||  |   |  | | ||| ||
Db       2949 TGAGCCACCGAGCCCGGCCAGACTTGACTTTTTTTTTTTTTTTT 2906


RESULT 19
US-09-949-016-15790
; Sequence 15790, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15790
;   LENGTH: 107827
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15790

  Query Match             10.2%;  Score 204.4;  DB 3;  Length 107827;
  Best Local Similarity   81.7%;  Pred. No. 3.3e-44;
  Matches  273;  Conservative    0;  Mismatches   56;  Indels    5;  Gaps    3;

Qy       1195 ATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTT 1254
             ||| | | ||  |||||     |  |||   |||| |||||||||| || ||||||||||
Db      74266 ATTGGTTCATTCTTTCTTTTGTTCTTTCTGTTTTTTTTTTTTTTGAGACCGAGTCTCACTC 74325

Qy       1255 TGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCA 1314
             |||||||||||||||||| |||||| | || ||||||||||||||||||||| |||||
Db      74326 TGTCACCCAGGCTGGAGTGCAGTGG-TATGATCTCGGCTCACTGCAACCTCTGCTTCCCA 74384

Qy       1315 GGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--A 1370
             |||||||||||||||| ||  ||||  |||| | ||||||||||||||  ||||||
Db      74385 GGTTCAAGCGATTCTCATGCCTCAGTCTCCCAAGTAGCTGGGATTACAGATGCATGCCAC 74444

Qy       1371 CCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
```

```
Db      74445 CAAGCCCAGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGTCAGGCT 74504

Qy       1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490

Db      74505 GGTCTCAAACTCCTGACCTCAAGTGATCCACCCCCCTCAGCCTCCCAAAGTGCTGGGATT 74564

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCT 1524

Db      74565 ACAGGTGTGAGCCACTGCGCCCAGCCTCTTTTTT 74598


RESULT 20
US-09-949-016-13747
; Sequence 13747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13747
;   LENGTH: 264665
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13747

  Query Match             10.2%;  Score 204.4;  DB 3;  Length 264665;
  Best Local Similarity   70.6%;  Pred. No. 5.1e-44;
  Matches  315;  Conservative    0;  Mismatches  126;  Indels    5;  Gaps    3;

Qy       1139 TTAGCCATGTGTGACAAAAGAGAGGCCATTTCAGTGTTGAAATTGTTCAGAGAAGTATTT 1198

Db     230063 TTAGGCAGATACAAAACAAAATTATTTCATGCATTCCTTTTGTTCTACCAATTAAATTTT 230122

Qy       1199 GATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTC 1258

Db     230123 GATTAGTACAATATTTGGTTCTTTTGTTTTTTTTTTTTTGAGACAAAGTCTCACTCTGTC 230182

Qy       1259 ACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTT 1318

Db     230183 ACCCAGTCTAGAGTGCTGTGGC-GTGATCTTGGCTCACTGCAACCTCCGCCTCCCAGGTT 230241

Qy       1319 CAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCA 1376

Db     230242 CAAGTGATTCTTCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGCGCGTGCCACCACA 230301

Qy       1377 --TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCC 1434

Db     230302 CCCGGCTAATTTTTGTATATTTTGTAGAGACAGGATTTTGCCATGTTGGCCAGGCTATTC 230361

Qy       1435 TTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAG 1494

Db     230362 GTGAACTCCTGACCTCAGGTGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATTACAG 230421

Qy       1495 GCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCA 1554

Db     230422 GTGTGAGCCACCGTGCCCAGCCGATTAGTACAGTATTTCTTTTCCAGCCTCTGTGCTTTC 230481

Qy       1555 TTTCTATCTTCTTATTCATTTTGGAA 1580

Db     230482 ATTTTTTTCTCTTCTTGTTTCTTGAA 230507


RESULT 21
US-09-949-016-16874/c
; Sequence 16874, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16874
;   LENGTH: 31867
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16874

  Query Match             10.2%;  Score 203.8;  DB 3;  Length 31867;
  Best Local Similarity   83.5%;  Pred. No. 2.7e-44;
  Matches  268;  Conservative    0;  Mismatches   47;  Indels    6;  Gaps    3;

Qy       1208 TTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCT 1267

Db      31472 TTCTCAAGTTTTTTTGTTGCTTTTTTGTTTTGAAAGGAGTCTCGCTCTGTCACCCAGGCT 31413

Qy       1268 GGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 1327

Db      31412 GGAGTGCAATGGC-GTGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 31354

Qy       1328 CTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG---CACCACCATGCCTA 1382

Db      31353 CTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCACCCGCCACCACGCCCGGCTAA 31294

Qy       1383 ATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTC 1442

Db      31293 TTTTTTGTATTTTTAGTAGAGAGGGGGTTTCACCATGTTGGCCAGGCTGGTCTGGAACTC 31234

Qy       1443 CTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGC 1502

Db      31233 CTGACCTCAGGTGATCCGCCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGC 31174

Qy       1503 CACCGTGCCCAGCCTGTTTTC 1523

Db      31173 CACTGTGCCCGGCCAATTCTC 31153


RESULT 22
US-09-949-016-12864/c
; Sequence 12864, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12864
;   LENGTH: 31875
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12864

  Query Match             10.2%;  Score 203.8;  DB 3;  Length 31875;
  Best Local Similarity   83.5%;  Pred. No. 2.7e-44;
  Matches  268;  Conservative    0;  Mismatches   47;  Indels    6;  Gaps    3;

Qy       1208 TTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCT 1267
             ||||||  | |||||  ||  | ||||  |||  ||  |||||||  || ||||||||||||
Db      31472 TTCTCAAGTTTTTTTGTTGCTTTTTTGTTTTGAAAGGAGTCTCGCTCTGTCACCCAGGCT 31413

Qy       1268 GGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 1327
             ||||| || |||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db      31412 GGAGTGCAATGGC-GTGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATT 31354

Qy       1328 CTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG---CACCACCATGCCTA 1382
             ||||||  |||||  ||||||  |||||||||||||||||  |  |   |  | ||  ||  |
Db      31353 CTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCACCCGCCACCACGCCCGGCTAA 31294

Qy       1383 ATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTC 1442
              ||||||||||||||||||||  | |||||  ||||||||  |||||||  | ||  ||||||
Db      31293 TTTTTTGTATTTTTAGTAGAGAGGGGGTTTCACCATGTTGGCCAGGCTGGTCTGGAACTC 31234

Qy       1443 CTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGC 1502
             |||||  ||||||||||  ||||||||||||||||||||  ||||||||||||||||||||
Db      31233 CTGACCTCAGGTGATCCGCCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGC 31174

Qy       1503 CACCGTGCCCAGCCTGTTTTC 1523
              ||| ||||||  |||  || ||
Db      31173 CACTGTGCCCGGCCAATTCTC 31153


RESULT 23
US-09-949-016-17593
; Sequence 17593, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17593
;   LENGTH: 64024
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17593

  Query Match             10.2%;  Score 203.6;  DB 3;  Length 64024;
  Best Local Similarity   78.2%;  Pred. No. 4.2e-44;
  Matches  283;  Conservative    0;  Mismatches   74;  Indels    5;  Gaps    3;

Qy       1191 AAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTC 1250
             || |||||   |  | ||| |    | ||||| ||||| ||||||||| ||||||||||
Db      13671 AATTATTTTTGTTAGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGATACAGAGTCTC 13730

Qy       1251 ACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCT 1310
              ||| |||| |||||||||||| || |||| | | ||| ||||||||||||||||||||
Db      13731 ACTCTGTCGCCCAGGCTGGAGTGCAATGGC-GCGATCTTGGCTCACTGCAACCTCTGCCT 13789

Qy       1311 CCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATG 1368
              ||  |||||||| |||||||  ||  ||||| ||| || ||||||||| |||||||||  |
Db      13790 CCTGGGTTCAAGTGATTCTCGTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGCGCGAG 13849

Qy       1369 CACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCA 1426
              || |  |   |  |||| |||||||||||| |||||||  | |||||  |||||||| |||
Db      13850 CAACCACGCCCAGCTAAGTTTTGTATTTTTGGTAGAGATGGGGTTTCACCATGTTGGCCA 13909

Qy       1427 GGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGG 1486
              |||| | || |||||||||||| |||||||||||||||||||||||||||  ||||||  ||||
Db      13910 GGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCAGCCTTCCAAAGTGCTGG 13969

Qy       1487 GATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGA 1546
              |||||||||||| || |||| ||||| |||   |  || |  |     |||  ||| | |
Db      13970 GATTACAGGCATTAGTCACCATGCCCGGCCGCCTAGCTAATTTTTAAAATTATTTTGTCA 14029

Qy       1547 AG 1548
              ||
Db      14030 AG 14031


RESULT 24
US-09-949-016-12758
; Sequence 12758, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12758
;   LENGTH: 88490
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12758

  Query Match             10.2%;  Score 203.4;  DB 3;  Length 88490;
  Best Local Similarity   78.7%;  Pred. No. 5.6e-44;
  Matches  281;  Conservative    0;  Mismatches   71;  Indels    5;  Gaps    3;

Qy       1218 TTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGT 1277
              ||  || |||| |||||||||| |  ||||||  ||| ||||||||||||||||| ||||
Db      50407 TTCATAATTTTTTTTTTTTTTGAGATGGAGTCTTGCTTCGTCACCCAGGCTGGAGTGCAGT 50466

Qy       1278 GGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TC 1335
              || || | ||| ||||||||||||||||||||||||||||||||||||||||||    |
Db      50467 GG-TGCGATCTTGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGCCGC 50525

Qy       1336 AGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTATT 1393
              ||| |||||| |||||||||||||||| || |||  |  || ||  |||| |||||||||
Db      50526 AGCCTCCCGAGTAGCTGGGATTACAGGTGCCTGCCACCACACGCGGCTAAGTTTTGTATT 50585

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
              ||| || ||||| |||||| |||||||| ||||| | | ||||||||||||| |||||
Db      50586 TTTTGTGGAGACGGAGTTTTGCCATGTTGCCCAGGGTGGTCTTGAACTCCTGACCTCAGG 50645

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
              |||||  |||||||| |||||||| ||  |||||||||||||||||||||||  |||
Db      50646 TGATCTGCCCACCTCGGCCTCCCAGAGTGCTGGGATTACAGGCATGAGCCACCGCACCCG 50705

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATT 1570
              ||||  |  ||    |  |  |  ||||  | ||   ||   | | || || ||| |
Db      50706 GCCTTATAACTTTATTTATAAACATGTTGATAAAAATTTATCTACAATATTTTTAAT 50762
```

```
RESULT 25
US-09-949-016-14222
; Sequence 14222, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14222
;   LENGTH: 88736
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-14222

  Query Match             10.2%;  Score 203.4;  DB 3;  Length 88736;
  Best Local Similarity   78.7%;  Pred. No. 5.6e-44;
  Matches  281;  Conservative    0;  Mismatches   71;  Indels    5;  Gaps    3;

Qy       1218 TTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGT 1277
             ||  || |||| |||||||||| |  |||||  ||| ||||||||||||||||| ||||
Db      50653 TTCATAATTTTTTTTTTTTTGAGATGGAGTCTTGCTTCGTCACCCAGGCTGGAGTGCAGT 50712

Qy       1278 GGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TC 1335
             || || | ||| ||||||||||||||||||||||||||||||||||||||||||    |
Db      50713 GG-TGCGATCTTGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTGCCGC 50771

Qy       1336 AGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTATT 1393
             ||| |||||| |||||||||||||| || |||  |  || ||  |||| |||||||||
Db      50772 AGCCTCCCGAGTAGCTGGGATTACAGGTGCCTGCCACCACACGCGGCTAAGTTTTGTATT 50831

Qy       1394 TTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGG 1453
             ||| || ||||| |||||| ||||||||| ||||| | | ||||||||||||||| |||||
Db      50832 TTTTGTGGAGACGGAGTTTTGCCATGTTGCCCAGGGTGGTCTTGAACTCCTGACCTCAGG 50891

Qy       1454 TGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCA 1513
             |||||  |||||||| |||||||| ||  |||||||||||||||||||||||  |||
Db      50892 TGATCTGCCCACCTCGGCCTCCCAGAGTGCTGGGATTACAGGCATGAGCCACCGCACCCG 50951

Qy       1514 GCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATT 1570
             ||||  |  ||    |  |  |  |||| | ||   ||   | | || || ||| |
Db      50952 GCCTTATAACTTTATTTATAAACATGTTGATAAAAATTTATCTACAATATTTTTAAT 51008


RESULT 26
US-09-949-016-11749/c
; Sequence 11749, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11749
;   LENGTH: 144596
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(144596)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11749

  Query Match             10.1%;  Score 202.8;  DB 3;  Length 144596;
  Best Local Similarity   78.5%;  Pred. No. 1e-43;
  Matches  281;  Conservative    0;  Mismatches   72;  Indels    5;  Gaps    3;

Qy       1182 TGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAA 1241
             ||   | | |  || ||  | |  ||||      | |||||||||| | |||||||| |
Db      97169 TGGCTAAATATTTACTTTTTAAATTTTTATTTTTATTTTTATTTTTTTATTTTTTGAGA 97110

Qy       1242 CAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAA 1301
              || ||||||| |||| |||||||||||||| ||||||| |  |||| |||||| |||||
Db      97109 TGGAATCTCACTGTGTCGCCCAGGCTGGAGTGCAGTGGC-GCAGTCTTGGCTCAGTGCAA 97051

Qy       1302 CCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTAC 1359
              |||||||||||||||||||| | |||||| ||  ||||| |||||| ||||||||||||
Db      97050 CCTCTGCCTCCCAGGTTCAAACAATTCTCATGCCTCAGCCTCCCGAGTAGCTGGGATTAC 96991

Qy       1360 AGGCGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCA 1417
              ||||||||||  |  ||  |  |||||||||||||||||||||||| || | ||||| |||
Db      96990 AGGCGCATGCCACCACACCCAGCTAATTTTTGTATTTTTAGTAGACACGGGGTTTCACCA 96931

Qy       1418 TGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCA 1477
              |||||  |||||| | |||||||||||||| ||| |||||| ||||||||| ||||||||
Db      96930 TGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCAAGTGATCTACCCACCTCGGCCTCCCA 96871

Qy       1478 AAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA 1535
              |||  |||||||||||||| |||||||| || |  ||||   || | |  |  ||||
Db      96870 AAGTGCTGGGATTACAGGCGTGAGCCACCATGACTGGCCTAAATTTTTACTTTTTGTA 96813


RESULT 27
US-09-949-016-13035/c
; Sequence 13035, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13035
;   LENGTH: 144596
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(144596)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13035

  Query Match             10.1%;  Score 202.8;  DB 3;  Length 144596;
  Best Local Similarity   78.5%;  Pred. No. 1e-43;
  Matches  281;  Conservative    0;  Mismatches   72;  Indels    5;  Gaps    3;
```

```
Qy       1182 TGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAA 1241
Db      97169 TGGCTAAATATTTACTTTTTAAATTTTTTATTTTTATTTTTATTTTTTTATTTTTGAGA 97110

Qy       1242 CAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAA 1301
Db      97109 TGGAATCTCACTGTGTCGCCCAGGCTGGAGTGCAGTGGC-GCAGTCTTGGCTCAGTGCAA 97051

Qy       1302 CCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTAC 1359
Db      97050 CCTCTGCCTCCCAGGTTCAAACAATTCTCATGCCTCAGCCTCCCGAGTAGCTGGGATTAC 96991

Qy       1360 AGGCGCATGCACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCA 1417
Db      96990 AGGCGCATGCCACCACACCCAGCTAATTTTTGTATTTTTAGTAGACACGGGGTTTCACCA 96931

Qy       1418 TGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCA 1477
Db      96930 TGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCAAGTGATCTACCCACCTCGGCCTCCCA 96871

Qy       1478 AAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTA 1535
Db      96870 AAGTGCTGGGATTACAGGCGTGAGCCACCATGACTGGCCTAAATTTTTACTTTTTGTA 96813


RESULT 28
US-09-949-016-15681/c
; Sequence 15681, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15681
;   LENGTH: 80269
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15681

  Query Match             10.1%;  Score 202.6;  DB 3;  Length 80269;
  Best Local Similarity   75.6%;  Pred. No. 8.8e-44;
  Matches  291;  Conservative    0;  Mismatches   89;  Indels    5;  Gaps    3;

Qy       1217 CTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAG 1276
Db      13610 CATTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCACCCAGGCTGGAGTGCAG 13551

Qy       1277 TGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--T 1334
Db      13550 TGGC-ATGACCTCGGCTCTCTACAACCTCTGCCTCCTGGGTTCAAGCGATTCTCCTGCCT 13492

Qy       1335 CAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTAT 1392
Db      13491 CAGCCTCCTAAGTAGCTGGGATTACAGCTGCCTGCCACCACGCCCGGCTGATTTTTGTAT 13432

Qy       1393 TTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAG 1452
Db      13431 TTTTAGTAGAAACAAGGTTTCACCATGTTGGCTAGGCTGGTCTCAAACTCCTGACCTCAA 13372

Qy       1453 GTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCC 1512
Db      13371 ATGATCCACCCGCCTCAGCCTCCCAAAGGGCTGGGATTACAGGCATGAGCCACCGTGCCC 13312

Qy       1513 AGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTCA 1572
Db      13311 AGCCAGCACTGACCATTTCTGACAGTCTCTATGGAGACTGAATATTTTGGTTTAAAAAAA 13252

Qy       1573 TTTTGGAAGTAGTACACCTAAGTAA 1597
Db      13251 AGTCTGCCTTAGTAATCTTAAGCAA 13227


RESULT 29
US-09-949-016-15870/c
; Sequence 15870, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15870
;   LENGTH: 45418
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15870

  Query Match             10.1%;  Score 202.4;  DB 3;  Length 45418;
  Best Local Similarity   77.7%;  Pred. No. 7.5e-44;
  Matches  283;  Conservative    0;  Mismatches   76;  Indels    5;  Gaps    3;

Qy       1192 AGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCA 1251
Db      15216 ATTTTCTGTTTTTTTTTTGTTTGTTTGTTTTTTGTTTTTTTTTTTTTTGAGACAGTCTCG 15157

Qy       1252 CTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTC 1311
Db      15156 CTCTGTCGCCCAGGCTGGAGTGCAGTGGC-GCAATCTTGGCTCACTGCAACCTCTGCCTC 15098

Qy       1312 CCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC 1369
Db      15097 CCGAGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCCCAC 15038

Qy       1370 ACCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAG 1427
Db      15037 CATCACGTTCAGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCGCCATGTTGGCTAG 14978

Qy       1428 GCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGG 1487
Db      14977 GCTGCTCTCGAACTTCTGACCTCAGGTGATCCACCCACCTCAGCCTCCCAAAGTGCAGGT 14918

Qy       1488 ATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAA 1547
Db      14917 ATTACAGGCGTTAGCCACTGCGCCCAGCCTCTTTGTTCTTATACATTAAATTTTCATGGA 14858

Qy       1548 GCCT 1551
Db      14857 CCGT 14854


RESULT 30
US-09-949-016-16238/c
; Sequence 16238, Application US/09949016
; Patent No. 6812339
```

```
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16238
;   LENGTH: 59828
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(59828)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16238

  Query Match             10.1%;  Score 202.4;  DB 3;  Length 59828;
  Best Local Similarity   78.7%;  Pred. No. 8.6e-44;
  Matches  280;  Conservative    0;  Mismatches   71;  Indels    5;  Gaps    3;

Qy       1174 GTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTT 1233
             ||   |||||     |||  |   || | ||  |||         ||  | ||||| ||||
Db      52769 GTGAGAATTGCGAGTAGATTTGAATGCTCATTGTTTGAACATCTTTCATTTTTTTTTTTT 52710

Qy       1234 TTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCT 1293
             |||||| || |||||||||| | ||||||||||||||| |||| | || | ||| ||||
Db      52709 TTTTGAGACCGAGTCTCACTCTATCACCCAGGCTGGAGTGCAGT-GTTGGGATCTTGGCT 52651

Qy       1294 CACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCT 1351
              ||||||| |||||||||| ||||||||| |||||||||  ||||| |||||| |||||
Db      52650 TACTGCAAACTCTGCCTCCCGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCGAGTAGCT 52591

Qy       1352 GGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAG 1409
             |||||||||||||  |||    | | |  |||||||||||||||||||| |||||| | | |
Db      52590 GGGATTACAGGCGTGTGCCACCAAGCCCGCCTAATTTTTGTATTTTTGGTAGAGGCGGGG 52531

Qy       1410 TTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCA 1469
              |||| |||||||| ||||||| | ||  |||||||||| ||| ||||||||||||||||
Db      52530 TTTCACCATGTTGGCCAGGCTAGTCTCTAACTCCTGACCTCACGTGATCCACCCACCTCG 52471

Qy       1470 GCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTCTC 1525
              || ||||||||  |||||||||||||||||||||||||||||||||||| |  | ||
Db      52470 GCTTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTTTCATATC 52415


RESULT 31
US-09-949-016-12257
; Sequence 12257, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12257
;   LENGTH: 129327
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(129327)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12257

  Query Match             10.1%;  Score 202.4;  DB 3;  Length 129327;
  Best Local Similarity   81.6%;  Pred. No. 1.3e-43;
  Matches  271;  Conservative    0;  Mismatches   56;  Indels    5;  Gaps    3;

Qy       1195 ATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTT 1254
              || |  | || |  |||    | ||||| ||||| |||||||||| ||||||||||||
Db      29960 ATCTAGTCATTTACTCTATTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTC 30019

Qy       1255 TGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCA 1314
              |||| ||  |||||||||| |||||| | || ||| || ||||||||||||  |||| ||
Db      30020 TGTCGCCTGGGCTGGAGTGCAGTGG-TATGATCTTGGGTCACTGCAACCTCCACCTCTCA 30078

Qy       1315 GGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--A 1370
               ||||||| ||||||||||  ||||| |||||| |||||||||||||||| || |||
Db      30079 AGTTCAAGTGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGTGCCTGCCAC 30138

Qy       1371 CCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
              |   | || |||||||||||| |||||||||||||| |||||||||||||| ||||| |
Db      30139 CATGCCTGGCTAATTTTTGTAATTTTAGTAGAGACAAAGTTTCGCCATGTTGGCCAGGGT 30198

Qy       1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490
              | || |||||||||||| |||||| |||||||||||||||||||| ||  ||||||||
Db      30199 GGTCTCGAACTCCTGACCTCAGGTAATCCACCCACCTCAGCCTCCCACAGTGCTGGGATT 30258

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTGTTTT 1522
              |||||  |||||||||||||||||  || |
Db      30259 ACAGGTGTGAGCCACCGTGCCCAGCCATTTGT 30290


RESULT 32
US-09-949-016-15368
; Sequence 15368, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15368
;   LENGTH: 129327
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(129327)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15368

  Query Match             10.1%;  Score 202.4;  DB 3;  Length 129327;
  Best Local Similarity   81.6%;  Pred. No. 1.3e-43;
  Matches  271;  Conservative    0;  Mismatches   56;  Indels    5;  Gaps    3;

Qy       1195 ATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTT 1254
```

```
Db      29960 ATCTAGTCATTTACTCTATTTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTC 30019

Qy       1255 TGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCA 1314
              |||| ||  |||||||||| |||||| | || ||| || ||||||||||||  |||| ||
Db      30020 TGTCGCCTGGGCTGGAGTGCAGTGG-TATGATCTTGGGTCACTGCAACCTCCACCTCTCA 30078

Qy       1315 GGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--A 1370
              ||||||| ||||||||||  ||||| |||||| ||||||||||||||||  || |||
Db      30079 AGTTCAAGTGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGTGCCTGCCAC 30138

Qy       1371 CCACCATGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCT 1430
              |    | || |||||||||||| |||||||||||||| |||||||||||||| ||||| |
Db      30139 CATGCCTGGCTAATTTTTGTAATTTTAGTAGAGACAAAGTTTCGCCATGTTGGCCAGGGT 30198

Qy       1431 TGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATT 1490
              | || |||||||||| |||||| |||||||||||||||||||||| ||  ||||||||
Db      30199 GGTCTCGAACTCCTGACCTCAGGTAATCCACCCACCTCAGCCTCCCACAGTGCTGGGATT 30258

Qy       1491 ACAGGCATGAGCCACCGTGCCCAGCCTGTTTT 1522
              |||||  ||||||||||||||||||  || |
Db      30259 ACAGGTGTGAGCCACCGTGCCCAGCCATTTGT 30290
```

RESULT 33
US-09-949-016-16585/c
; Sequence 16585, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16585
; LENGTH: 36532
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(36532)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16585

Query Match 10.1%; Score 201.6; DB 3; Length 36532;
Best Local Similarity 82.0%; Pred. No. 1.1e-43;
Matches 269; Conservative 0; Mismatches 54; Indels 5; Gaps 3;

```
Qy       1206 TTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGG 1265
              ||||  |  |||  || |  |||  || |||||||  |||||||||  || ||||  ||||||
Db      28215 TTTTTTTTGATTGTTGTTGTTTGTTTGTTTTTGAGACAGAGTCTTGCTCTGTCGCCCAGG 28156

Qy       1266 CTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGA 1325
              ||||||||||||||  || ||| |||||||||||||||||||||||  ||||||||||
Db      28155 CTGGAGTACAGTGGC-ATGATCTTGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGA 28097

Qy       1326 TTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCT 1381
              ||||||||  |||||  |||||| ||||||||||||||||||||  |||   |   | || ||
Db      28096 TTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCGCCTGCCACCACGCCTGGCT 28037

Qy       1382 AATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACT 1441
              ||||||||||||||||||||||  |||||||  ||||||||  |||||||  | |||||||
Db      28036 AATTTTTGTATTTTTAGTAGAGATGGAGTTTCACCATGTTGGCCAGGCTGGTCTTGAACT 27977
```

```
Qy       1442 CCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAG 1501
              |||||| |||||||||||||| |||  |||||| ||||   |||||||||| ||||||
Db      27976 CCTGACCTCAGGTGATCCACCCGCCTTGGCCTCCAAAAGTGGTGGGATTACAGACATGAG 27917

Qy       1502 CCACCGTGCCCAGCCTGTTTTCTCAGAT 1529
              |||||  |||| |||  | |  |||| |
Db      27916 CCACCACGCCCGGCCAATATCTTCAGTT 27889
```

RESULT 34
US-09-949-016-16939/c
; Sequence 16939, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16939
; LENGTH: 67620
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16939

Query Match 10.1%; Score 201.6; DB 3; Length 67620;
Best Local Similarity 84.3%; Pred. No. 1.5e-43;
Matches 263; Conservative 0; Mismatches 44; Indels 5; Gaps 3;

```
Qy       1217 CTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAG 1276
              || ||| || || |||||||||| |||||||||||| |||| ||||||||||||||||||
Db      24013 CTGTTTTTTGTTTTTTTTTTTGAGACAGAGTCTCACTGTGTCTCCCAGGCTGGAGTACAG 23954

Qy       1277 TGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--T 1334
              ||| ||   |||||| ||||||||||||||||||  ||||||||||||||||||  |
Db      23953 TGG-TGCAATCTCGGTTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTCCTGCCT 23895

Qy       1335 CAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTAT 1392
              |||| ||| || |||||||||||||||| ||||||   |   |   ||||||||||||
Db      23894 CAGCCTCCTGAGTAGCTGGGATTACAGGTGCATGCCATCATGCCCAGCTAATTTTTGTAT 23835

Qy       1393 TTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAG 1452
              |||||||||||||| | ||||| |||||||| ||||||| | ||  ||||||||||  |||
Db      23834 TTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCCCAG 23775

Qy       1453 GTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCC 1512
              |||||||||||||||||||||||||||  |||||||||||||| |||||||| |||||
Db      23774 ATGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCATGCCC 23715

Qy       1513 AGCCTGTTTTCT 1524
              |||||  | | |
Db      23714 AGCCTCCTGTGT 23703
```

RESULT 35
US-09-949-016-17521/c
; Sequence 17521, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016

```
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17521
;   LENGTH: 87734
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17521

  Query Match            10.1%;  Score 201.6;  DB 3;  Length 87734;
  Best Local Similarity   81.0%;  Pred. No. 1.7e-43;
  Matches  272;  Conservative    0;  Mismatches   59;  Indels    5;  Gaps    3;

Qy       1206 TTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGG 1265
             | || |||| |  |||| ||| | |||||||||| |||||||||| || || ||||||||
Db      26852 TCTTTTCAGTTTCTTTTTTTTGTTTTTTTTTTGAGACAGAGTCTCGCTCTGCCACCCAGG 26793

Qy       1266 CTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGA 1325
             ||||||| || |||| ||| ||||||||| |||||| |||  ||||||||||||||||||
Db      26792 CTGGAGTGCAATGGC-GTGATCTCGGCTCTCTGCAATCTCCACCTCCCAGGTTCAAGCGA 26734

Qy       1326 TTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCAT--GCCT 1381
             ||||||||  ||| | |||||| ||||| ||| ||||||| | |||  |  |||  | ||
Db      26733 TTCTCCTGCCTCAACCTCCCGAGTAGCTCGGACTACAGGCACGTGCCACCACATCCGGCT 26674

Qy       1382 AATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACT 1441
             ||||||||||||||||||||||||| |||||  ||| |||   |||||  | || |||||
Db      26673 AATTTTTGTATTTTTAGTAGAGACAGGGTTTCAACATATTGGTGAGGCTGGTCTCGAACT 26614

Qy       1442 CCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAG 1501
             ||||||  |||||||||||||||  ||||||||||||||  ||||||||||||  |||
Db      26613 CCTGACCTCAGGTGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAA 26554

Qy       1502 CCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATT 1537
             ||||| ||||  |||| || || ||| | || || |
Db      26553 CCACCATGCCAGGCCTCTTCTCACAGTTTCTCTAGT 26518


RESULT 36
US-09-949-016-13469/c
; Sequence 13469, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13469
;   LENGTH: 46347
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13469

  Query Match            10.1%;  Score 201.4;  DB 3;  Length 46347;
  Best Local Similarity   85.0%;  Pred. No. 1.4e-43;
  Matches  261;  Conservative    0;  Mismatches   41;  Indels    5;  Gaps    3;

Qy       1216 TCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACA 1275
             ||||| | |||||  |||||||||| | |||||  | ||||||||||| |||||||| ||
Db      35745 TCTTTCTTTTTTTTTATTTTTTTGAGATAGAGTTTTGCTTTGTCACCCATGCTGGAGTGCA 35686

Qy       1276 GTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG-- 1333
             |||||   | ||| |||||||||||||||||||||||||||||||  |||||||||
Db      35685 GTGGC-ACGATCTTGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAATGATTCTCCTGCC 35627

Qy       1334 TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCACCATGC--CTAATTTTTGTA 1391
             |||||||||||| |||||||||||||||||| |||  |  ||  |  ||||||||||||
Db      35626 TCAGCTTCCCGAGTAGCTGGGATTACAGGCGCCTGCCACCACACCCAGCTAATTTTTGTA 35567

Qy       1392 TTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCA 1451
             |||||||||||||||| |||||  |||||||| ||| | | | ||||||||||||||| |||
Db      35566 TTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAAGTTGGTCTTGAACTCCTGACCTCA 35507

Qy       1452 GGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511
             ||||||||||  |||||||| ||||||  |||||||||||||  |||||||||| |||
Db      35506 GGTGATCCACCTGCCTCAGCCTTCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGCGCC 35447

Qy       1512 CAGCCTG 1518
             |||||||
Db      35446 CAGCCTG 35440


RESULT 37
US-09-949-016-15830/c
; Sequence 15830, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15830
;   LENGTH: 192506
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15830

  Query Match            10.1%;  Score 201.4;  DB 3;  Length 192506;
  Best Local Similarity   73.0%;  Pred. No. 2.8e-43;
  Matches  300;  Conservative    0;  Mismatches  106;  Indels    5;  Gaps    3;

Qy       1169 TCAGTGTTGAAATTGTTCAGAGAAGTATTTGATTATGTTTTCTCAGATCTTTTTATTTTT 1228
             |||  ||| | ||| |   |   | | ||   ||   ||||||    | ||||| |||||
Db      76912 TCAAGGTTTATATTTTGGTGGTCATTTTTCTTTTTCTTTTTCTGTTTTTTTTTTTTTTTT 76853

Qy       1229 ATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGGCTGGAGTACAGTGGC-TGTGGTC 1287
              |||||||||| |||||||||||| |||||||||||| ||||    | |     ||
Db      76852 TTTTTTTTTGAGACAGAGTCTCACTCTGTCACCCAGGCTTGAGTGTGCAGTCGCACAATC 76793

Qy       1288 TCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGA 1345
              || |||||||| ||||||  |||| |||||||||||| |||||||||  |||| ||| ||
Db      76792 TCAGCTCACTGTAACCTCCACCTCTCAGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGA 76733

Qy       1346 ATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCTAATTTTTGTATTTTTAGTAGAG 1403
              ||||||||||| | |||||||||    |  | || |||||||||||| ||||||| ||
Db      76732 GTAGCTGGGATTATAAGCGCATGCTGCCACGCCTGGCTAATTTTTGTAATTTTAGTAAAG 76673

Qy       1404 ACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACTCCTGACTTCAGGTGATCCACCC 1463
              |||| ||||| |||||||| ||||||| | |||||||||||| | ||||||||||||||
Db      76672 ACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGGCCTCAGGTGATCCACCC 76613
```

```
Qy       1464 ACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCCCAGCCTGTTTTC 1523
Db      76612 ACCTTGGTTTCCCAAAGTGCTGGGATTACAGGTATGAGCTACTGCACCCAGCCATTTTTC 76553

Qy       1524 TCAGATCCTGTATTTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTCATT 1574
Db      76552 ATTGCTAATGAGTTTTTCTCCTAATTTTCTCCTTTAATATGTCTAGTTATT 76502


RESULT 38
US-09-949-016-169667
; Sequence 169667, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 169667
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-169667

  Query Match             10.1%;  Score 201.2;  DB 3;  Length 601;
  Best Local Similarity   81.1%;  Pred. No. 2e-44;
  Matches  271;  Conservative    0;  Mismatches   58;  Indels    5;  Gaps    3;

Qy       1193 GTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCAC 1252
Db        255 GTATTTTCTTTTGGTTTCTTTCCTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTTGC 314

Qy       1253 TTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCC 1312
Db        315 TCTGTCACCCAGGCTGGAGTGCAGT-GATGTGATCTCGGCTTACTGCAACCTCTGCCTCC 373

Qy       1313 CAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCA 1370
Db        374 CGGGTTCAAGCGATTCTCTTGCCTCAGCCTCCCAAGTAGCTGGGACTAAAGGTGTGTACC 433

Qy       1371 CCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGG 1428
Db        434 ACCACACCCCACTAATTTTTGTATTTTTAGTAGAAACGGGGTTTCACCATGTTGGCCAGG 493

Qy       1429 CTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGA 1488
Db        494 CTGGTCTCGAACTCCTGACCTCAGATGACCCACCCACCTCAGCCTCCCAAAGTGCTGGGA 553

Qy       1489 TTACAGGCATGAGCCACCGTGCCCAGCCTGTTTT 1522
Db        554 TTACAGGCGTGAGCCACCGCGCCCAGCCTCTTTT 587


RESULT 39
US-09-949-016-171021
; Sequence 171021, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 171021
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-171021

  Query Match             10.1%;  Score 201.2;  DB 3;  Length 601;
  Best Local Similarity   81.7%;  Pred. No. 2e-44;
  Matches  268;  Conservative    1;  Mismatches   54;  Indels    5;  Gaps    3;

Qy       1206 TTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTGTCACCCAGG 1265
Db        163 TTTTTTTTGATTGTTGTTGTTTGTTTGTTTTTGAGACAGAGTCTTGCTCTGTCGCCCAGG 222

Qy       1266 CTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCGA 1325
Db        223 CTGGAGTACAGTGGC-ATGATCTTGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGA 281

Qy       1326 TTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGC--ACCACCATGCCT 1381
Db        282 TTCTCCTGCCTCAGCCTCCYGAGTAGCTGGGATTACAGGCGCCTGCCACCACGCCTGGCT 341

Qy       1382 AATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCTTGAACT 1441
Db        342 AATTTTTGTATTTTTAGTAGAGATGGAGTTTCACCATGTTGGCCAGGCTGGTCTTGAACT 401

Qy       1442 CCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAG 1501
Db        402 CCTGACCTCAGGTGATCCACCCGCCTTGGCCTCCAAAAGTGGTGGGATTACAGACATGAG 461

Qy       1502 CCACCGTGCCCAGCCTGTTTTCTCAGAT 1529
Db        462 CCACCACGCCCGGCCAATATCTTCAGTT 489


RESULT 40
US-09-949-016-11990/c
; Sequence 11990, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11990
;   LENGTH: 99748
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-11990

  Query Match             10.1%;  Score 201.2;  DB 3;  Length 99748;
  Best Local Similarity   81.1%;  Pred. No. 2.3e-43;
  Matches  271;  Conservative    0;  Mismatches   58;  Indels    5;  Gaps    3;
```

```
Qy       1193 GTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCAC 1252
Db      14250 GTATTTTCTTTTGGTTTCTTTCCTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTTGC 14191

Qy       1253 TTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCC 1312
Db      14190 TCTGTCACCCAGGCTGGAGTGCAGT-GATGTGATCTCGGCTTACTGCAACCTCTGCCTCC 14132

Qy       1313 CAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCA 1370
Db      14131 CGGGTTCAAGCGATTCTCTTGCCTCAGCCTCCCAAGTAGCTGGGACTAAAGGTGTGTACC 14072

Qy       1371 CCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGG 1428
Db      14071 ACCACACCCCACTAATTTTTGTATTTTTAGTAGAAACGGGGTTTCACCATGTTGGCCAGG 14012

Qy       1429 CTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGA 1488
Db      14011 CTGGTCTCGAACTCCTGACCTCAGATGACCCACCCACCTCAGCCTCCCAAAGTGCTGGGA 13952

Qy       1489 TTACAGGCATGAGCCACCGTGCCCAGCCTGTTTT 1522
Db      13951 TTACAGGCGTGAGCCACCGCGCCCAGCCTCTTTT 13918
```

RESULT 41
US-09-949-016-16518/c
; Sequence 16518, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16518
; LENGTH: 99749
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16518

Query Match 10.1%; Score 201.2; DB 3; Length 99749;
Best Local Similarity 81.1%; Pred. No. 2.3e-43;
Matches 271; Conservative 0; Mismatches 58; Indels 5; Gaps 3;

```
Qy       1193 GTATTTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCAC 1252
Db      14250 GTATTTTCTTTTGGTTTCTTTCCTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTTGC 14191

Qy       1253 TTTGTCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCC 1312
Db      14190 TCTGTCACCCAGGCTGGAGTGCAGT-GATGTGATCTCGGCTTACTGCAACCTCTGCCTCC 14132

Qy       1313 CAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCA 1370
Db      14131 CGGGTTCAAGCGATTCTCTTGCCTCAGCCTCCCAAGTAGCTGGGACTAAAGGTGTGTACC 14072

Qy       1371 CCACCATGC--CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGG 1428
Db      14071 ACCACACCCCACTAATTTTTGTATTTTTAGTAGAAACGGGGTTTCACCATGTTGGCCAGG 14012

Qy       1429 CTTGCCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGA 1488
Db      14011 CTGGTCTCGAACTCCTGACCTCAGATGACCCACCCACCTCAGCCTCCCAAAGTGCTGGGA 13952

Qy       1489 TTACAGGCATGAGCCACCGTGCCCAGCCTGTTTT 1522
Db      13951 TTACAGGCGTGAGCCACCGCGCCCAGCCTCTTTT 13918
```

RESULT 42
US-09-834-700-13
; Sequence 13, Application US/09834700
; Patent No. 6958214
; GENERAL INFORMATION:
; APPLICANT: Braun, A.
; TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: 24736-2035
; CURRENT APPLICATION NUMBER: US/09/834,700
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/217,251
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: 60/240,335
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 162025
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: allele
; LOCATION: 83587
; OTHER INFORMATION: Nucleotide sequence of Chromosome 17 with genomic
; OTHER INFORMATION: sequence of the allelic variant AKAP10-6
US-09-834-700-13

Query Match 10.1%; Score 201.2; DB 3; Length 162025;
Best Local Similarity 66.7%; Pred. No. 2.9e-43;
Matches 348; Conservative 0; Mismatches 168; Indels 6; Gaps 4;

```
Qy       1197 TTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTTGAAACAGAGTCTCACTTTG 1256
Db      56490 TGGTTTATGATTTGGAGTTTTTTGGTTTTTGTTTTTTGTTTTTGACAGAGTCTTACTCTG 56549

Qy       1257 TCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 1316
Db      56550 TCACCCAGGCTGGAGAGCAGTGGC-ATAATCTCAGCTCACTGCAACCTCTGTCTCCCAGG 56608

Qy       1317 TTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC 1374
Db      56609 TTTGAGCGATTCTCCTGCCTCAGCCTACTGAGTAGCTGGGATTACAGGCGCGTGCCACCA 56668

Qy       1375 CA--TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTG 1432
Db      56669 CACCCGGCTAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGGTGG 56728

Qy       1433 CCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTAC 1492
Db      56729 TCTTGAACTCCTGACCTCAGGTAATCCACCTGCCTCAGCCTCCCAAAGTGCTGAGATTAC 56788

Qy       1493 AGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTT 1552
Db      56789 AGGCGTGAGCCACCGCGCCTGGCATGGTTTGGAGTTTTAATCTGTAGTTTTAATAAAGAT 56848

Qy       1553 CATTTCTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCA 1612
Db      56849 AGTGCTTATGTTTGTGTTTCTTATATTTCTTGGTACTCTTGGGTA-ATTTGTAAGATCCC 56907

Qy       1613 AATATCTTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAG 1672
Db      56908 CATATCTACACAAGAAGTCCATTTTCAATTCTTTTCTTCAGACTGTTTATTTTATTTTAT 56967

Qy       1673 CTGATGTTATGTTTCTTTCAAATTATTCATTTCTCTATCTCA 1714
```

```
Db      56968 TTTATTTTATTTTTATGTTTGAGATGGAGTCTCGCTGTGTCA 57009


RESULT 43
US-09-834-700-14
; Sequence 14, Application US/09834700
; Patent No. 6958214
; GENERAL INFORMATION:
;  APPLICANT: Braun, A.
;  TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND
;  TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING THE SAME
;  FILE REFERENCE: 24736-2035
;  CURRENT APPLICATION NUMBER: US/09/834,700
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: 60/217,251
;  PRIOR FILING DATE: 2000-07-10
;  PRIOR APPLICATION NUMBER: 60/240,335
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 162025
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 129600
;   OTHER INFORMATION: Nucleotide sequence of chromosome 17 containing
;   OTHER INFORMATION: the genomic sequence of the allelic variant
;   OTHER INFORMATION: AKAP10-7
US-09-834-700-14

  Query Match             10.1%;  Score 201.2;  DB 3;  Length 162025;
  Best Local Similarity   66.7%;  Pred. No. 2.9e-43;
  Matches  348;  Conservative    0;  Mismatches  168;  Indels    6;  Gaps    4;

Qy       1197 TTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTG 1256
             | | ||||| |||     | ||   | ||| | |||| |||   |||||||||| ||| ||
Db      56490 TGGTTTATGATTTGGAGTTTTTTGGTTTTTGTTTTTTGTTTTTGACAGAGTCTTACTCTG 56549

Qy       1257 TCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 1316
             ||||||||||||||||  |||||||   |  |||| |||||||||||||||||| ||||||||
Db      56550 TCACCCAGGCTGGAGAGCAGTGGC-ATAATCTCAGCTCACTGCAACCTCTGTCTCCCAGG 56608

Qy       1317 TTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC 1374
             ||  |||||||||||||  ||||| | | || ||||||||||||||||||| |||  |
Db      56609 TTTGAGCGATTCTCCTGCCTCAGCCTACTGAGTAGCTGGGATTACAGGCGCGTGCCACCA 56668

Qy       1375 CA--TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTG 1432
              ||   | ||||||||||||||||||||||||  | ||||| |||||||| ||||| | |
Db      56669 CACCCGGCTAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGGTGG 56728

Qy       1433 CCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTAC 1492
              |||||||||||||| |||||| |||||||  |||||||||||||||  ||| ||||||
Db      56729 TCTTGAACTCCTGACCTCAGGTAATCCACCTGCCTCAGCCTCCCAAAGTGCTGAGATTAC 56788

Qy       1493 AGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTT 1552
              |||| |||||||||| |||  || || |||      |      | ||||    |   |
Db      56789 AGGCGTGAGCCACCGCGCCTGGCATGGTTTGGAGTTTAATCTGTAGTTTTAATAAAGAT 56848

Qy       1553 CATTTCTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCA 1612
                |  ||| ||  | ||  || |      | |      ||  |   | ||| ||| | |
Db      56849 AGTGCTTATGTTTGTGTTTCTTATATTTCTTGGTACTCTTGGGTA-ATTTGTAAGATCCC 56907

Qy       1613 AATATCTTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAG 1672
              ||||||    || |   ||   |||  ||||| |    | |    |   || | | ||
Db      56908 CATATCTACACAAGAAGTCCATTTTCAATTCTTTTCTTCAGACTGTTTATTTTATTTTAT 56967

Qy       1673 CTGATGTTATGTTTCTTTCAAATTATTCATTTCTCTATCTCA 1714
              | || |||| ||| | |   |       | || || | |||
Db      56968 TTTATTTTATTTTTATGTTTGAGATGGAGTCTCGCTGTGTCA 57009


RESULT 44
US-09-834-700-17
; Sequence 17, Application US/09834700
; Patent No. 6958214
; GENERAL INFORMATION:
;  APPLICANT: Braun, A.
;  TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND
;  TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING THE SAME
;  FILE REFERENCE: 24736-2035
;  CURRENT APPLICATION NUMBER: US/09/834,700
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: 60/217,251
;  PRIOR FILING DATE: 2000-07-10
;  PRIOR APPLICATION NUMBER: 60/240,335
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 162025
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Nucleotide sequence of chromosome 17 containing
;   OTHER INFORMATION: the genomic sequence of the predominate allele of
;   OTHER INFORMATION: the AKAP10 gene
US-09-834-700-17

  Query Match             10.1%;  Score 201.2;  DB 3;  Length 162025;
  Best Local Similarity   66.7%;  Pred. No. 2.9e-43;
  Matches  348;  Conservative    0;  Mismatches  168;  Indels    6;  Gaps    4;

Qy       1197 TTGATTATGTTTTCTCAGATCTTTTTATTTTTATTTTTTTTGAAACAGAGTCTCACTTTG 1256
             | | ||||| |||     | ||   | ||| | |||| |||   |||||||||| ||| ||
Db      56490 TGGTTTATGATTTGGAGTTTTTTGGTTTTTGTTTTTTGTTTTTGACAGAGTCTTACTCTG 56549

Qy       1257 TCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 1316
             ||||||||||||||||  |||||||   |  |||| |||||||||||||||||| ||||||||
Db      56550 TCACCCAGGCTGGAGAGCAGTGGC-ATAATCTCAGCTCACTGCAACCTCTGTCTCCCAGG 56608

Qy       1317 TTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC 1374
             ||  |||||||||||||  ||||| | | || ||||||||||||||||||| |||  |
Db      56609 TTTGAGCGATTCTCCTGCCTCAGCCTACTGAGTAGCTGGGATTACAGGCGCGTGCCACCA 56668

Qy       1375 CA--TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTG 1432
              ||   | ||||||||||||||||||||||||  | ||||| |||||||| ||||| | |
Db      56669 CACCCGGCTAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGGTGG 56728

Qy       1433 CCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTAC 1492
              |||||||||||||| |||||| |||||||  |||||||||||||||  ||| ||||||
Db      56729 TCTTGAACTCCTGACCTCAGGTAATCCACCTGCCTCAGCCTCCCAAAGTGCTGAGATTAC 56788

Qy       1493 AGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTT 1552
              |||| |||||||||| |||  || || |||      |      | ||||    |   |
Db      56789 AGGCGTGAGCCACCGCGCCTGGCATGGTTTGGAGTTTAATCTGTAGTTTTAATAAAGAT 56848

Qy       1553 CATTTCTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCA 1612
                |  ||| ||  | ||  || |      | |      ||  |   | ||| ||| | |
Db      56849 AGTGCTTATGTTTGTGTTTCTTATATTTCTTGGTACTCTTGGGTA-ATTTGTAAGATCCC 56907

Qy       1613 AATATCTTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAG 1672
              ||||||    || |   ||   |||  ||||| |    | |    |   || | | ||
Db      56908 CATATCTACACAAGAAGTCCATTTTCAATTCTTTTCTTCAGACTGTTTATTTTATTTTAT 56967

Qy       1673 CTGATGTTATGTTTCTTTCAAATTATTCATTTCTCTATCTCA 1714
              | || |||| ||| | |   |       | || || | |||
Db      56968 TTTATTTTATTTTTATGTTTGAGATGGAGTCTCGCTGTGTCA 57009


RESULT 45
US-09-834-700-18
; Sequence 18, Application US/09834700
```

```
; Patent No. 6958214
; GENERAL INFORMATION:
;  APPLICANT: Braun, A.
;  TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND
;  TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING THE SAME
;  FILE REFERENCE: 24736-2035
;  CURRENT APPLICATION NUMBER: US/09/834,700
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: 60/217,251
;  PRIOR FILING DATE: 2000-07-10
;  PRIOR APPLICATION NUMBER: 60/240,335
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
;   LENGTH: 162025
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 156,277
;   OTHER INFORMATION: Nucleotide sequence of chromosome 17 containing
;   OTHER INFORMATION: the genomic sequence of the allelic variant
;   OTHER INFORMATION: AKAP10-1
US-09-834-700-18

  Query Match             10.1%;  Score 201.2;  DB 3;  Length 162025;
  Best Local Similarity   66.7%;  Pred. No. 2.9e-43;
  Matches  348;  Conservative    0;  Mismatches  168;  Indels    6;  Gaps    4;

Qy       1197 TTGATTATGTTTTCTCAGATCTTTTTATTTTATTTTTTTTGAAACAGAGTCTCACTTTG 1256
              | | ||||| |||       | ||  | ||| | |||| |||   |||||||| ||| ||
Db      56490 TGGTTTATGATTTGGAGTTTTTTGGTTTTTGTTTTTTGTTTTTGACAGAGTCTTACTCTG 56549

Qy       1257 TCACCCAGGCTGGAGTACAGTGGCTGTGGTCTCGGCTCACTGCAACCTCTGCCTCCCAGG 1316
              ||||||||||||||  |||||||  |  |||| |||||||||||||||||| ||||||||
Db      56550 TCACCCAGGCTGGAGAGCAGTGGC-ATAATCTCAGCTCACTGCAACCTCTGTCTCCCAGG 56608

Qy       1317 TTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGGATTACAGGCGCATGCACCAC 1374
              ||  ||||||||||||  ||||| | | || ||||||||||||||||||| |||  |
Db      56609 TTTGAGCGATTCTCCTGCCTCAGCCTACTGAGTAGCTGGGATTACAGGCGCGTGCCACCA 56668

Qy       1375 CA--TGCCTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCCATGTTGACCAGGCTTG 1432
              ||   | |||||||||||||||||||||||  | ||||| |||||||| |||| | |
Db      56669 CACCCGGCTAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGGTGG 56728

Qy       1433 CCTTGAACTCCTGACTTCAGGTGATCCACCCACCTCAGCCTCCCAAAGCACTGGGATTAC 1492
              ||||||||||||| |||||| ||||||| |||||||||||||||  ||| ||||||
Db      56729 TCTTGAACTCCTGACCTCAGGTAATCCACCTGCCTCAGCCTCCCAAAGTGCTGAGATTAC 56788

Qy       1493 AGGCATGAGCCACCGTGCCCAGCCTGTTTTCTCAGATCCTGTATTTGTTTCTGAAGCCTT 1552
              |||| |||||||||| |||  || || |||       |       | ||||    |    |
Db      56789 AGGCGTGAGCCACCGCGCCTGGCATGGTTTGGAGTTTTAATCTGTAGTTTTAATAAAGAT 56848

Qy       1553 CATTTCTATCTTCTTATTCATTTTGGAAGTAGTACACCTAAGTAAGGTTTTTAACAATCA 1612
                |   ||| ||  | ||  || |      | |      ||  |  |  ||| ||| |  |
Db      56849 AGTGCTTATGTTTGTGTTTCTTATATTTCTTGGTACTCTTGGGTA-ATTTGTAAGATCCC 56907

Qy       1613 AATATCTTTGGAAAATTCCCTGGTTCCTTTCTTATTCCTACAAAAATATGTTCAGTATAG 1672
              ||||||    || |   ||   |||  ||||| |    | |    |   || | | ||
Db      56908 CATATCTACACAAGAAGTCCATTTTCAATTCTTTTCTTCAGACTGTTTATTTTATTTTAT 56967

Qy       1673 CTGATGTTATGTTTCTTTCAAATTATTCATTTCTCTATCTCA 1714
                | || |||| ||| | |   |       | || || | |||
Db      56968 TTTATTTTATTTTTATGTTTGAGATGGAGTCTCGCTGTGTCA 57009


Search completed: January 22, 2006, 21:30:33
Job time : 257.224 secs
```

(upside down text)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:10:03 ; Search time 5419.54 Seconds
(without alignments)
17274.699 Million cell updates/sec

Title: US-09-728-552A-3_COPY_39000_41000
Perfect score: 2001
Sequence: 1 tgatccaccagccttggcct.....ctaattcagtgaacacttca 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	417.2	20.8	494	9	AQ598756	AQ598756 HS_5338_B
c	2	214.4	10.7	744	9	BZ610799	BZ610799 WHADJ52TR
	3	212.8	10.6	871	9	BZ603854	BZ603854 WHADC93TF
	4	211.4	10.6	658	7	CN266085	CN266085 170004249
c	5	208.8	10.4	443	2	BF991881	BF991881 PM4-GN001
c	6	208.2	10.4	435	2	BF991882	BF991882 PM4-GN001
c	7	206.8	10.3	643	1	AW975010	AW975010 EST387115
	8	206.8	10.3	683	1	AV703063	AV703063 AV703063
c	9	206.4	10.3	361	1	AW572721	AW572721 hc86b11.x
	10	206.4	10.3	866	9	AQ889099	AQ889099 HS_5263_A
	11	206	10.3	605	9	AQ347764	AQ347764 RPCI11-13
	12	206	10.3	812	2	BG026977	BG026977 602294004
	13	205.6	10.3	710	5	BU849843	BU849843 AGENCOURT
c	14	205.4	10.3	544	1	AU147414	AU147414 AU147414
	15	205.2	10.3	2768	4	CR936687	CR936687 Homo sapi
	16	205	10.2	455	9	AQ037381	AQ037381 CIT-HSP-2
	17	205	10.2	462	6	CD699526	CD699526 EST16162
	18	205	10.2	693	9	BZ612424	BZ612424 WHADM75TR
	19	205	10.2	704	2	BE787136	BE787136 601476688
c	20	204.4	10.2	555	5	BX484164	BX484164 DKFZp686J
	21	204.4	10.2	829	9	BZ599806	BZ599806 WHADX62TR
c	22	204.2	10.2	458	1	AI733856	AI733856 zo19c03.y
c	23	203.6	10.2	390	1	AA808875	AA808875 nt86h02.s
c	24	203.6	10.2	438	7	CR739570	CR739570 CR739570
	25	203.4	10.2	545	2	BF854170	BF854170 MR2-EN009
	26	203	10.1	526	11	CR961104	CR961104 Homo sapi
c	27	203	10.1	3165	4	CR859576	CR859576 Pongo pyg
	28	202.8	10.1	492	6	CD701193	CD701193 EST17717
c	29	202.8	10.1	888	5	BU957691	BU957691 AGENCOURT
c	30	202.6	10.1	742	9	AQ582332	AQ582332 RPCI-11-4
	31	202.4	10.1	470	10	AG190706	AG190706 Pan trogl
c	32	202.2	10.1	449	9	AQ769842	AQ769842 HS_3170_B
	33	202.2	10.1	466	9	AQ822778	AQ822778 HS_2205_B
c	34	202.2	10.1	552	1	AI479068	AI479068 tm30g04.x
c	35	202.2	10.1	552	1	AL712937	AL712937 DKFZp686B
c	36	202.2	10.1	930	5	BU517124	BU517124 AGENCOURT
	37	202	10.1	935	2	BF968880	BF968880 602270854
	38	201.8	10.1	439	1	AL712324	AL712324 DKFZp686E
	39	201.8	10.1	662	10	AG037584	AG037584 Pan trogl
	40	201.8	10.1	678	1	AV729893	AV729893 AV729893
c	41	201.8	10.1	1102	3	BM458632	BM458632 AGENCOURT
	42	201.6	10.1	456	1	AW408767	AW408767 UI-HF-BM0
c	43	201.6	10.1	524	1	AI791156	AI791156 ab53e09.x
	44	201.6	10.1	714	2	BG545785	BG545785 602573059
	45	201.6	10.1	3181	4	CR936826	CR936826 Homo sapi
c	46	201.2	10.1	711	10	AG010148	AG010148 Homo sapi
	47	201.2	10.1	730	6	CB308016	CB308016 AGENCOURT
c	48	201	10.0	749	6	CB962143	CB962143 AGENCOURT
	49	200.8	10.0	692	10	AG173021	AG173021 Pan trogl
c	50	200.8	10.0	696	5	BU616112	BU616112 UI-H-DF0-
c	51	200.6	10.0	439	1	AL707781	AL707781 DKFZp686J
c	52	200.6	10.0	618	6	CD579489	CD579489 EST_PSF00
c	53	200.6	10.0	618	6	CD579508	CD579508 EST_PSF00
c	54	200.6	10.0	861	6	CD517715	CD517715 AGENCOURT
c	55	200.6	10.0	907	6	CD558895	CD558895 AGENCOURT
c	56	200.4	10.0	490	1	AI792575	AI792575 qi76f07.y
	57	200.4	10.0	690	10	AG182694	AG182694 Pan trogl
c	58	200.4	10.0	739	1	AW957154	AW957154 EST369224
c	59	200.4	10.0	917	5	BU153359	BU153359 AGENCOURT
	60	200.2	10.0	548	1	AW962006	AW962006 EST374079
c	61	200.2	10.0	671	1	AL048002	AL048002 DKFZp586D
c	62	200.2	10.0	679	7	CN265762	CN265762 170004247
c	63	200.2	10.0	2213	4	CR623485	CR623485 full-leng
	64	200	10.0	597	1	AV733437	AV733437 AV733437
	65	200	10.0	598	9	B92813	B92813 CIT-HSP-216
	66	200	10.0	693	10	AG179363	AG179363 Pan trogl
	67	199.8	10.0	588	9	AQ581201	AQ581201 RPCI-11-4
	68	199.8	10.0	613	9	BZ610480	BZ610480 WHADK14TF
	69	199.8	10.0	624	2	BF854308	BF854308 MR2-EN009
c	70	199.8	10.0	670	3	BM671184	BM671184 UI-E-CQ1-
	71	199.8	10.0	681	9	AQ543621	AQ543621 RPCI-11-3
c	72	199.8	10.0	909	5	BQ881589	BQ881589 AGENCOURT
	73	199.6	10.0	528	2	BG005696	BG005696 RC1-GN019
c	74	199.6	10.0	624	5	BX508814	BX508814 DKFZp686O
c	75	199.4	10.0	501	5	BX485916	BX485916 DKFZp686E
	76	199.4	10.0	561	6	CD688632	CD688632 EST5154 h
c	77	199.4	10.0	712	6	CF146965	CF146965 UI-HF-CB0
c	78	199.4	10.0	849	8	DN602222	DN602222 HESC3_78_
	79	199.4	10.0	933	5	BQ959144	BQ959144 AGENCOURT
c	80	199.2	10.0	554	7	CN270785	CN270785 170005999
c	81	199.2	10.0	664	10	AG127255	AG127255 Pan trogl
	82	199.2	10.0	687	10	AG085195	AG085195 Pan trogl
c	83	199.2	10.0	826	5	BQ438665	BQ438665 AGENCOURT
c	84	199.2	10.0	870	9	AQ739907	AQ739907 HS_5506_A
	85	199.2	10.0	924	9	AQ459187	AQ459187 HS_5078_A
c	86	199	9.9	516	9	AQ112451	AQ112451 CIT-HSP-2
	87	199	9.9	541	2	BG682496	BG682496 602623935
	88	199	9.9	548	1	AI583291	AI583291 tt56g02.x
	89	199	9.9	626	9	AQ485636	AQ485636 RPCI-11-2
c	90	199	9.9	740	10	CL983267	CL983267 HSC 01080
	91	199	9.9	945	5	BU527620	BU527620 AGENCOURT
c	92	198.8	9.9	329	1	AI185394	AI185394 qe52e01.x
c	93	198.8	9.9	392	1	AL045476	AL045476 DKFZp434H
	94	198.8	9.9	394	9	AQ605998	AQ605998 HS_5383_A
c	95	198.8	9.9	521	10	CG734335	CG734335 GS1-174H8

```
    96   198.8     9.9     587  9  AQ345662                    AQ345662 RPCI11-10
    97   198.8     9.9     600  9  AQ622755                    AQ622755 HS_5339_B
c   98   198.8     9.9     707  11  CR974543                    CR974543 Homo sapi
c   99   198.8     9.9     844  2  BF694941                    BF694941 602080953
   100   198.6     9.9     392  9  AZ254675                    AZ254675 UP_561-7J
```

ALIGNMENTS

```
RESULT 1
AQ598756
LOCUS       AQ598756                494 bp    DNA     linear   GSS 08-JUN-1999
DEFINITION  HS_5338_B1_D03_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=914 Col=5 Row=H, genomic survey sequence.
ACCESSION   AQ598756
VERSION     AQ598756.1  GI:5029968
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 914  row: H  column: 5
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=914 Col=5 Row=H"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match           20.8%;  Score 417.2;  DB 9;  Length 494;
  Best Local Similarity   96.6%;  Pred. No. 1.4e-45;
  Matches  425;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1062 AGTTAGCACCTTGGTATTTTCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCA 1121
              ||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 AGTTATCACCTTGGTATTTTCTTTATCCTTCATTGCAATACAGAAGATGCTTCATATGCA 60

Qy       1122 CAGTGGTGTCAGGTCACATCAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAAC 1181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         61 CAGTGGTGTCAGGTCACATCAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAAC 120

Qy       1182 CGGAAAGGAAAGGCACCCATTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGC 1241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        121 CGGAAAGGAAAGGCACCCATTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTANAGAGC 180

Qy       1242 AGGCAGATGCTTTGAAGGCGTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGC 1301
              | |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 ACGCAGATGCTTTGAAGGCGTGGTAACACAAGTCTTCATTAATCTCCACGCAGGACTTGC 240

Qy       1302 ACTTCTACTATGCCTAGGCTGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCC 1361
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        241 ACTTCTACTATGCCTATGCTGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCC 300

Qy       1362 CTAACTAACTGAAGCCAGGTGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTC 1421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAACTAACTGAAGCCAGGTGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTC 360

Qy       1422 TTGAAATCTCCCACTTCTTTCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATT 1481
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        361 TTGAAATCTCCCACTTCTTTCTACGTCAGAAGAGCCAAGCTGATTATTNTAGTTGGAATT 420

Qy       1482 TAGAAATTTTTAAAAATTAT 1501
                || |||  ||| ||| |
Db        421 AGAAATTTTACAAATATTCT 440


RESULT 2
BZ610799/c
LOCUS       BZ610799                744 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHADJ52TR Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-21I8, genomic survey sequence.
ACCESSION   BZ610799
VERSION     BZ610799.1  GI:31519360
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-21I8"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN

  Query Match           10.7%;  Score 214.4;  DB 9;  Length 744;
```

```
  Best Local Similarity   73.0%;  Pred. No. 6.2e-19;
  Matches  305;  Conservative    0;  Mismatches  106;  Indels    7;  Gaps    2;

Qy        268 TGTTTGATATGAAATAAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTA 327
              || |||| | || | |    | |||||||||  |        ||| || ||| ||  | |
Db        610 TGATTGAGAAGATAGATATTTATCAACAAGTCTTCAAACAAATTTAAAATTAAAATGAGG 551

Qy        328 TTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTT 387
               |   | | |  |    ||  |||   | ||  | |   |   || |||  ||||   ||
Db        550 ATACATTCTATGAGAGCATTTAATAGCAGAAGGTGACTTAGGAAAGTTAGAAAAGGCTTT 491

Qy        388 TGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGT 447
                  ||||| ||||||| |||  ||| |||||||||| ||||||||||||  |||
Db        490 ----CCTGGCCATGCGCAGTGGCTTATGCTTGTAATCCCAGCACTTTGGGAGGCTGAGGC 435

Qy        448 GGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGT 507
              ||| ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db        434 GGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCTGT 375

Qy        508 CTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTAC 567
              ||||||  |||||  |||||||||||  ||| ||||||||  |||||||| ||||| ||||
Db        374 CTCTACCAAAAATACAAAAATTAGCTGGGTATGGTGGCAGGCACCTGTAATCCCAGCTAC 315

Qy        568 TCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGA 627
                 |||||  |||||  |||||||||||||  |||  ||||||  |||||||||||||| | ||
Db        314 CTGGGAGGCTGAGGCAGGAGAATTGCTTAAACTCAGGAGGCGGAGGTTGCAGTAAGCTGA 255

Qy        628 GAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||    | |||||||||||||||||||  ||| ||||||||||  ||     | || ||||
Db        254 GATCACGCCCCTGCACTCCAGCCTGGCGACACAGCAAGACTCTATCTCAAAAAAAAAA 197


RESULT 3
BZ603854
LOCUS       BZ603854                871 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHADC93TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-19P17, genomic survey sequence.
ACCESSION   BZ603854
VERSION     BZ603854.1  GI:31512316
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 871)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .871
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-19P17"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN

  Query Match            10.6%;  Score 212.8;  DB 9;  Length 871;
  Best Local Similarity   77.7%;  Pred. No. 9.5e-19;
  Matches  283;  Conservative    0;  Mismatches   77;  Indels    4;  Gaps    2;

Qy        344 AATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCG 403
              | |  ||| | |   |  ||||||| |   |   | || ||||    || ||||||||||
Db        462 ATTGCAATGTGAGCGAGAAATACACCTTGTTGTGTTAAAAGCTGCTCAAAGGGCCAGGCG 521

Qy        404 CAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAG 463
              |||| | ||   |||||||||| | ||||||||||||| | |||||||||||||||||
Db        522 CAGTGGATCACACCTGTAATCCTAGCACTTTGGGAGGCCGAAGTGGGCGGATCACCTGAG 581

Qy        464 GTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCA 523
              ||| ||||||| |||||||||||||||||||||||||||    |||||||||||||| ||
Db        582 GTCGGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACTGCATCTCTACTAAAAACACA 641

Qy        524 AAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGA 583
              |||||||||||||| |||||||   |||||||||  |||| ||||||||||||| ||||| |
Db        642 AAAATTAGCCAGGCGTGGTGGTGAGCACCTGTAACCCCAGCTACTCAGGAGGCTGAGGCA 701

Qy        584 GGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCAC 640
              ||||||  |||||  |||    |||| |||||  ||||||  | || |||    || |||||||
Db        702 AGAGAATCGCTTGGACCCCAGAGGCGGAGGGTGCAGTGAGCCCAGATTGTGCCACTGCAT 761

Qy        641 TCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAA 699
              |||||||| ||||| || |||||||||| | || | || |||   |  ||||||| || ||
Db        762 TCCAGCCTGGGCAAGAGGGCAAGACTCCGTCAAAAAAAAAAAAGCCGCTGAAATTTTGGAA 821

Qy        700 ATGA 703
                ||
Db        822 TGGA 825


RESULT 4
CN266085
LOCUS       CN266085                658 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000424984279 GRN_ES Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN266085
VERSION     CN266085.1  GI:47282499
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 658)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
   PUBMED   15146197
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 658   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .658
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, cell lines H1, H7, and
                     H9"
                     /clone_lib="GRN_ES"
```

/note="oligo dT primed, full-length enriched cDNA library from undifferentiated hES cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions"

ORIGIN

Query Match 10.6%; Score 211.4; DB 7; Length 658;
Best Local Similarity 75.1%; Pred. No. 1.6e-18;
Matches 277; Conservative 0; Mismatches 91; Indels 1; Gaps 1;

```
Qy        389 GAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTG 448
              ||||  |||||||||| || |||  ||||||||||||| || |||||| ||| |||
Db        168 GAAAAAGGCCAGGCGCGGTGGCTTATGCCTGTAATCCCAGCATTTTGGGAAGCCGAGGCA 227

Qy        449 GGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTC 508
              |||  |||||||||||| |||||||| |||||||||| |||||||||||||||||| |||
Db        228 GGCAAATCACCTGAGGTTAGGAGTTTGAGACCAGCCTGACCAACATGGTGAAACCCCGTC 287

Qy        509 TCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACT 568
              ||||||||||  ||||||||||  |||||| ||| |  |||||| |||||| |||||
Db        288 TCTACTAAAAATACAAAAATTAGCTGGGTGTGATGGTACACACCTGTGGTCCCAGCTACT 347

Qy        569 CAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG 628
              | ||||| ||||| |||||||| |||||||||| | ||| ||||||||||| | | |||
Db        348 CGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGAAGGCGGAGGTTGCAGTGAGCTGAG 407

Qy        629 AT-GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTT 687
              ||   || || |||||||||||||| ||||||| ||||||| |    | || ||||
Db        408 ATCACCATTGTACTCCAGCCTGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAA 467

Qy        688 GAAATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGAAATTCATCTTTGTTCATTTAT 747
               ||| |||  |  || ||  ||    ||   |  || |  ||||| ||  | ||
Db        468 AAAAGTGTAGACGTGGGTGCTAGTGGACTGTGTGATAGTACATCATCATTAATTATGAGG 527

Qy        748 TTTTACTTG 756
              || |   ||
Db        528 TTCTTTCTG 536
```

RESULT 5
BF991881/c
LOCUS BF991881 443 bp mRNA linear EST 23-JAN-2001
DEFINITION PM4-GN0017-291000-004-b03 GN0017 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF991881
VERSION BF991881.1 GI:12398204
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 443)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM4&t2=PM4-GN0017-291000-004-b03&t3=2000-10-29&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 20
High quality sequence stop: 326.

FEATURES Location/Qualifiers
source 1. .443
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="GN0017"
/note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 10.4%; Score 208.8; DB 2; Length 443;
Best Local Similarity 81.0%; Pred. No. 4.1e-18;
Matches 256; Conservative 0; Mismatches 57; Indels 3; Gaps 1;

```
Qy        369 TTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAA 428
              | ||| | | |||       ||     |||  |||   || |||| ||||||||||||||
Db        414 TCAAACTCTAAAAAGAGGCTGGGTGTGGCTGGGCATGGTGGCTCATGCCTGTAATCCCAG 355

Qy        429 CACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGC 488
              ||||||||||||||||||| || |||||||||||||||||||| ||||||||||||
Db        354 CACTTTGGGAGGCCAAGGTGCGCAGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGC 295

Qy        489 CAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATG 548
              |||||||||||||||||||||||||||||  ||||||||||  || ||||||||
Db        294 CAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGCGCA 235

Qy        549 CACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGT 608
              ||||||  |||||  ||||||  |||||  |||||  ||||||||  |||||||||  ||||  |
Db        234 TGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCAGGGAGTT 175

Qy        609 GGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACT 665
              ||||||||||| | |||||||     |||||||||||||||||||||| |||||||||
Db        174 GGAGGTTGCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGCAACACAGCAAGACT 115

Qy        666 CCATAAAGACAACAAA 681
              || |  | | |  |||
Db        114 CCGTGTAAAAACAAAA 99
```

RESULT 6
BF991882/c
LOCUS BF991882 435 bp mRNA linear EST 23-JAN-2001
DEFINITION PM4-GN0017-291000-004-b04 GN0017 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF991882
VERSION BF991882.1 GI:12398205
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 435)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics

```
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM4&t2=PM4-GN0017-
            291000-004-b04&t3=2000-10-29&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 15
            High quality sequence stop: 78.
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="GN0017"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1:
                     SmaI; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
ORIGIN

  Query Match             10.4%;  Score 208.2;  DB 2;  Length 435;
  Best Local Similarity   78.2%;  Pred. No. 4.9e-18;
  Matches  262;  Conservative    0;  Mismatches   70;  Indels    3;  Gaps    1;

Qy        350 ATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAG 409
              || ||||       |   |   ||| | | |||       ||     |||  |||   || |
Db        427 ATCTTAATCGAAGACTTATCAAAACTCTAAAAANAGGCTGGGTGTGGCTGGGCATGGTGG 368

Qy        410 CTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGG 469
              ||| ||||||||||| || ||||||||||||||||||||| || ||||||||||||||||||
Db        367 CTCATGCCTGTAATCNCAGCACTTTGGGAGGCCAAGGTGCGCAGATCACCTGAGGTCAGG 308

Qy        470 AGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATT 529
              ||||| |||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db        307 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATT 248

Qy        530 AGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAA 589
              |||  || |||||||||      |||||| ||||| |||||| ||||| |||||||||
Db        247 AGCTGGGCGTGGTGGCGCATGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAA 188

Qy        590 TTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGC 646
              | |||||||||  |||| ||||||||||||| | |||||||     ||||||||||||||
Db        187 TCGCTTGAACCAGGGAGTTGGAGGTTGCAGTGAGCCGAGATCACACCACTGCACTCCAGC 128

Qy        647 CTGGCAACAGAGCAAGACTCCATAAAGACAACAAA 681
              ||||||||| |||||||||| |  | | |  |||
Db        127 CTGGCAACACAGCAAGACTCCGTGTAAAAACAAAA 93


RESULT 7
AW975010/c
LOCUS       AW975010                643 bp    mRNA    linear   EST 02-JUN-2000
DEFINITION  EST387115 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW975010
VERSION     AW975010.1  GI:8166213
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C.,
            Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 343
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1. .643
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGN"
                     /note="Vector: pBluescriptSKm"
ORIGIN

  Query Match             10.3%;  Score 206.8;  DB 1;  Length 643;
  Best Local Similarity   80.9%;  Pred. No. 6.5e-18;
  Matches  254;  Conservative    0;  Mismatches   57;  Indels    3;  Gaps    1;

Qy        372 AATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACAC 431
              ||| |  ||| |  ||||   | ||||||||||| || ||||  || ||||||||||| |||
Db        426 AATGAAAAAAAATATTTGCCCTTGGCCAGGCGCGGTGGCTCACGCTTGTAATCCCAGCAC 367

Qy        432 TTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAA 491
              |||||||||| |||||||| |||||||||||||   |||||||| ||||||||||||||||
Db        366 TTTGGGAGGCAAAGGTGGGTGGATCACCTGAGGCTAGGAGTTTGAGACCAGCCTGGCCAA 307

Qy        492 CATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCAC 551
              |||||||||||||  | ||||||||||  ||||||||||  ||||||||||  ||||||
Db        306 CATGGTGAAACCCCATATCTACTAAAAATACAAAAATTAGCGGGGTGTGGTGGTATGCAC 247

Qy        552 CTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGA 611
              ||||| ||||| |||||| ||||| |||||  |||||||||||||||| |||||| |||
Db        246 CTGTAATCCCACCTACTCCGGAGGCTGAGGCTGGAGAATTGCTTGAACCCAGGAGGGGGA 187

Qy        612 GGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCA 668
              ||||||||| | | |||||   | || ||||||||||||||  |  |||  | |||  |
Db        186 GGTTGCAGTGAGCTGAGATCACGCCATTGCACTCCAGCCTGGGTAACGAGAGAAACTGTA 127

Qy        669 TAAAGACAACAAAA 682
              |    | ||  ||||
Db        126 TTTCAAAAAAAAAA 113


RESULT 8
AV703063
LOCUS       AV703063                683 bp    mRNA    linear   EST 08-OCT-2000
DEFINITION  AV703063 ADB Homo sapiens cDNA clone ADBBYB10 5', mRNA sequence.
ACCESSION   AV703063
VERSION     AV703063.1  GI:10719393
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G.,
            Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z.,
            Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S.,
            Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA ADB clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
```

```
           Tel: 86-21-50801919(ex.45)
           Fax: 86-21-50801922
           Email: hanzg@chgc.sh.cn
           This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ADBBYB10"
                     /tissue_type="Adrenal gland"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="ADB"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             10.3%;  Score 206.8;  DB 1;  Length 683;
  Best Local Similarity   84.2%;  Pred. No. 6.4e-18;
  Matches  245;  Conservative    0;  Mismatches   43;  Indels    3;  Gaps    1;

Qy        395 GGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGA 454
              |||||||||| || |||| ||||| ||||| || |||||||| ||||  |||  ||| ||
Db        295 GGCCAGGCGCGGTGGCTCATGCCTCTAATCTCAGCACTTTGGAAGGCTGAGGCAGGCAGA 354

Qy        455 TCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACT 514
              |||||| |||||||||||| ||||||| |||||||||||||||||||| |||||||||
Db        355 TCACCTAAGGTCAGGAGTTCAAGACCATCCTGGCCAACATGGTGAAACCATGTCTCTACT 414

Qy        515 AAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAG 574
              |||||  |||||||||||||| |||||||   ||||||||||||| ||||||||||
Db        415 AAAAATACAAAAATTAGCCAGGAGTGGTGGTGCGCACCTGTAGTCCCAGCTACTCAGGAG 474

Qy        575 GTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---G 631
              | ||||| |||||||| ||||||| | | ||||||||||||||||| | |||||||   |
Db        475 GCTGAGGCAGGAGAATCGCTTGAATCCATGAGGTGGAGGTTGCAGTGAGCCGAGATCGAG 534

Qy        632 TCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
               ||||||| |||||||||| |||||||||||||   |    | || || |
Db        535 CCACTGCAATCCAGCCTGGCGACAGAGCAAGACTCTGTCTCAAAAAAAANA 585


RESULT 9
AW572721/c
LOCUS       AW572721                361 bp    mRNA    linear   EST 13-MAR-2000
DEFINITION  hc86b11.x2 NCI_CGAP_Kid1 Homo sapiens cDNA clone IMAGE:2899581
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AW572721
VERSION     AW572721.1  GI:7237454
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
            Seq primer: -40UP from Gibco
            High quality sequence stop: 342.
FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2899581"
                     /tissue_type="kidney"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Kid1"
                     /note="Vector: pAMP10; mRNA made from invasive kidney
                     tumor, cDNA made by oligo-dT priming.  Non-directionally
                     cloned.  Size-selected on agarose gel, average insert
                     size 600 bp. Reference:  Krizman et al. (1996) Cancer
                     Research 56:5380-5383."
ORIGIN

  Query Match             10.3%;  Score 206.4;  DB 1;  Length 361;
  Best Local Similarity   83.4%;  Pred. No. 9.1e-18;
  Matches  247;  Conservative    0;  Mismatches   46;  Indels    3;  Gaps    1;

Qy        390 AAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGG 449
              | |  ||||||| || || ||||   |||| ||||||| ||||||||||||||||||||
Db        319 ATAGAGGCCAGGTGCGGTGGCTCACACCTGCAATCCCAGCACTTTGGGAGGCCAAGGTGG 260

Qy        450 GCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCT 509
              || |||||||||||||||||||||  |||||||||||||||||||||||||| |||||||
Db        259 GCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAAGCCTGTCT 200

Qy        510 CTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTC 569
              | |||||||| |||||||||||||| ||  ||||||   |||||||| ||||| ||||||
Db        199 CAACTAAAAATGCAAAAATTAGCCGGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTC 140

Qy        570 AGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG- 628
              |||||| ||||| |||||||||||||||||||| ||||| || |||||||||| | |||||
Db        139 AGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGTGGTTGCAGTGAGCCGAGA 80

Qy        629 --ATGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
                ||| ||||| |||||||||||||| ||||| ||| ||  |       || ||||
Db         79 TCATGCCACTGTACTCCAGCCTGGCAATAGAGCTAGATTCTGTCTCAGAAAAAAAA 24


RESULT 10
AQ889099
LOCUS       AQ889099                866 bp    DNA     linear   GSS 10-NOV-1999
DEFINITION  HS_5263_A2_A05_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=9031 Col=10 Row=A, genomic survey sequence.
ACCESSION   AQ889099
VERSION     AQ889099.1  GI:6345289
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 866)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
```

```
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 9031  row: A  column: 10
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 866.
FEATURES              Location/Qualifiers
     source           1. .866
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /clone="Plate=9031 Col=10 Row=A"
                      /sex="male"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             10.3%;  Score 206.4;  DB 9;  Length 866;
  Best Local Similarity   70.4%;  Pred. No. 6.6e-18;
  Matches  307;  Conservative    0;  Mismatches  121;  Indels    8;  Gaps    2;

Qy        352 TTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCT 411
              |||  ||  | |   | | |    |  || |  | |    | |||||||||  || |||
Db        269 TTTCCAATGTCACTTATATGATGGCTATAAAAAGTATACTTTAGGCCAGGCGTGGTGGCT 328

Qy        412 CCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAG 471
              | ||||||||||||||| ||||||||||||||||||  || ||||||||||| |||||||
Db        329 CATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAGGTGGATCACCTGAAGTCAGGAG 388

Qy        472 TTTAAGACCAG-----CCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAA 526
              || ||||||||     |||||||||||||||||||||||| |||| ||||||||  |||||
Db        389 TTCAAGACCAGACCACCCTGGCCAACATGGTGAAACCCTATCTCCACTAAAAATACAAAA 448

Qy        527 ATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGA 586
              ||||||  ||| ||||||    |||||||||||||| |||||  ||||| ||| | ||||
Db        449 ATTAGCTGGGTATGGTGGTGCACACCTGTAGTCCCAGCTACTTGGGAGGCTGAAGCAGGA 508

Qy        587 GAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCC 643
              |||| |||||||||| |||||  |||||||||||| | ||||||   || ||||||||||
Db        509 GAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTCAGCCGAGATTGTGCCACTGCACTCC 568

Qy        644 AGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGA 703
              |||||||| ||||||  |||||| |    | || ||||       ||| ||||   | |
Db        569 AGCCTGGCGACAGAGGGAGACTCCGTCTCAAGAAAAAAAAAAAAAAAAAAAGTGTGCTTTA 628

Qy        704 GTTGTACCTATCTTCATTTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAG 763
              | | |  ||      ||  | |||||     ||  |||  | | || |||   || |
Db        629 ATAGAAGGTAAGAGGCTTGCAAAAATTAGAAATTTCTCAAGTCTCTTAACTGACCAGGGA 688

Qy        764 AGCTTCCAGCAATTTT 779
              | ||| |   ||  ||
Db        689 AACTTACCTGAAACTT 704


RESULT 11
AQ347764
LOCUS       AQ347764                 605 bp    DNA     linear   GSS 07-MAY-1999
DEFINITION  RPCI11-137M1.TV RPCI-11 Homo sapiens genomic clone RPCI-11-137M1,
            genomic survey sequence.
ACCESSION   AQ347764
VERSION     AQ347764.1  GI:4172660
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .605
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="GDB:7552512"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-137M1"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /clone_lib="RPCI-11"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             10.3%;  Score 206;  DB 9;  Length 605;
  Best Local Similarity   82.8%;  Pred. No. 8.5e-18;
  Matches  260;  Conservative    0;  Mismatches   50;  Indels    4;  Gaps    2;

Qy        373 ATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACT 432
              | |||  |||| |   |||||||||||||||||| ||||  | |||||||||| ||||
Db         31 AGTATATAAGAAGTGCTAAATGGGCCAGGCGCAGTGGCTCACGTCTGTAATCCCAGCACT 90

Qy        433 TTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAAC 492
              |||||||||| ||||||| |||||||||||||||||| ||| |||||||||||| |||||
Db         91 TTGGGAGGCCGAGGTGGGTGGATCACCTGAGGTCAGGGGTTCAAGACCAGCCTGACCAAC 150

Qy        493 ATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACC 552
              ||||||||||  ||||||||||||  ||||||||||||||  || ||||| |  ||
Db        151 ATGGTGAAACCTCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGATGGCACGTGCC 210

Qy        553 TGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAG 612
              |||||||||| |||||| ||||| ||||| | |||||| |||||||| ||||||||||
Db        211 TGTAGTCCCAGCTACTCGGGAGGCTGAGGCATGAGAATCGCTTGAACCCAGGAGGTGGAG 270

Qy        613 GTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCA 668
              |||||||| | |||||    || |||||||||||||||| || |||||| | |||||||
Db        271 GTTGCAGTGAGCCGAGGTTGTGCCACTGCACTCCAGCCTGGGTAACAGAACGAGACTCCG 330

Qy        669 TAAAGACAACAAAA 682
              |   || || || |
Db        331 TTTCGAAAAAAAGA 344


RESULT 12
BG026977
LOCUS       BG026977                 812 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  602294004F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4388761 5',
            mRNA sequence.
ACCESSION   BG026977
VERSION     BG026977.1  GI:12415140
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 812)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10075  row: g  column: 02
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1. .812
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4388761"
                     /tissue_type="osteosarcoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_86"
                     /note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.533 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             10.3%;  Score 206;  DB 2;  Length 812;
  Best Local Similarity   85.0%;  Pred. No. 7.6e-18;
  Matches  243;  Conservative    0;  Mismatches   40;  Indels    3;  Gaps    1;

Qy        400 GGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACC 459
              || ||||| |||| |||||||||||||| |||||| |||||||||||||||| |||||||
Db        242 GGTGCAGTGGCTCATGCCTGTAATCCCAGCACTTTAGGAGGCCAAGGTGGGCAGATCACC 301

Qy        460 TGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAA 519
              ||||||||||||||  |||||||||||||| |||||||||||||||| ||||||||||||
Db        302 TGAGGTCAGGAGTTCGAGACCAGCCTGGCAAACATGGTGAAACCCCGTCTCTACTAAAAG 361

Qy        520 CGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGA 579
              ||||||||||  |||||||||||| || |||||| ||||| || |||||||||| |||
Db        362 TACAAAAATTAGCTGGGTGTGGTGGCAAGCGCCTGTAATCCCAGCTGCTCAGGAGGCTGA 421

Qy        580 GGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGATGTC---ACT 636
              || | |||||| |||||||| |||||| |||||||||||| | || |||| |   |||
Db        422 GGCACGAGAATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCAAGATTACACTACT 481

Qy        637 GCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              ||||||||||||||||||||||  |||||| ||    | || ||||
Db        482 GCACTCCAGCCTGGCAACAGAGTGAGACTCTATCTCAAAAAAAAAA 527


RESULT 13
BU849843
LOCUS       BU849843                710 bp    mRNA    linear   EST 16-OCT-2002
DEFINITION  AGENCOURT_10440376 NIH_MGC_109 Homo sapiens cDNA clone
            IMAGE:6598537 5', mRNA sequence.
ACCESSION   BU849843
VERSION     BU849843.1  GI:24034806
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2831  row: i  column: 01
            High quality sequence stop: 592.
FEATURES             Location/Qualifiers
     source          1. .710
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6598537"
                     /tissue_type="teratocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_109"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: EcoRI; Site_2:
                     XhoI; cDNA made by oligo-dT priming.  Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G).  Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             10.3%;  Score 205.6;  DB 5;  Length 710;
  Best Local Similarity   75.7%;  Pred. No. 9.1e-18;
  Matches  281;  Conservative    0;  Mismatches   85;  Indels    5;  Gaps    2;

Qy        317 TTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTA 376
              || | |  |   |   || || ||    |  || | ||| |  | | || |||  ||
Db        333 TTCAGGTATATATATATATAAGTAAAGGAATTAAGTAAAAATAAAATTTCAGTTACTTTT 392

Qy        377 TTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGG 436
              | ||||  | |  ||   |||| |  || || |||| ||||||||||||| ||||||||
Db        393 TAAAAGCACCTGAAATCTGGCCGGATGCGGTGGCTCATGCCTGTAATCCCACCACTTTGG 452

Qy        437 GAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGG 496
              |||||| ||| |||| ||||||||||||||||||||| ||||||||||||||||||||||
Db        453 GAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGG 512

Qy        497 TGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTA 556
              ||||||||  ||||||||||||  |||||||||||| || |||||| || || ||||||
Db        513 TGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGTCAGGCGCCTGTA 572

Qy        557 GTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTG 616
              |||||| |||||||||||| ||||| |||||||| ||||||||||  ||||  |||||||
Db        573 ATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGNGAGGCAGAGGTTG 632

Qy        617 CAGTAACCCGAGATGTCAC---TGCACTCCAGCCT--GGCAACAGAGCAAGACTCCATAA 671
              |||| | | |||||  |||   || ||||||||||  || |  ||||| |||||||||
Db        633 CAGTGAGCTGAGATTGCACCATTGTACTCCAGCCTGNGGGACAAGAGCGAGACTCCATCT 692

Qy        672 AGACAACAAAA 682
              | || ||||
Db        693 CAAAAAAAAAA 703


RESULT 14
AU147414/c
LOCUS       AU147414                544 bp    mRNA    linear   EST 05-AUG-2002
DEFINITION  AU147414 MAMMA1 Homo sapiens cDNA clone MAMMA1000605 3', mRNA
            sequence.
ACCESSION   AU147414
VERSION     AU147414.1  GI:11008935
KEYWORDS    EST.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T.,
            Sugano,S. and Isogai,T.).
  TITLE     HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
            Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M.,
            Nakamura,Y., Nagai,T., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MAMMA1000605"
                     /tissue_type="mammary gland"
                     /clone_lib="MAMMA1"
                     /note="Vector: pME18SFL3"
ORIGIN

  Query Match             10.3%;  Score 205.4;  DB 1;  Length 544;
  Best Local Similarity   85.7%;  Pred. No. 1.1e-17;
  Matches  240;  Conservative    0;  Mismatches   37;  Indels    3;  Gaps    1;

Qy        390 AAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGG 449
              | |  ||||||| || || ||||   |||| ||||||| ||||||||||||||||||||
Db        288 ATAGAGGCCAGGTGCGGTGGCTCACACCTGCAATCCCAGCACTTTGGGAGGCCAAGGTGG 229

Qy        450 GCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCT 509
              || ||||||||||||||||||||||  |||||||||||||||||||||||||||| |||||||
Db        228 GCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAAGCCTGTCT 169

Qy        510 CTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTC 569
              |||||||||| |||||||||||||| ||  ||||||   || |||||| ||||| ||||||
Db        168 CTACTAAAAATGCAAAAATTAGCCGGGCATGGTGGTGGGCNCCTGTAATCCCAGCTACTC 109

Qy        570 AGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG- 628
              |||||| ||||| |||||||||||||||||||| ||||| || |||||||||| | |||||
Db        108 AGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGTGGTTGCAGTGAGCCGAGA 49

Qy        629 --ATGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTC 666
                ||| ||||| |||||||||||||||| ||||| ||| ||
Db         48 TCATGCCACTGTACTCCAGCCTGGCAATAGAGCTAGATTC 9


RESULT 15
CR936687
LOCUS       CR936687               2768 bp    mRNA    linear   HTC 23-FEB-2005
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686H23233 (from clone DKFZp686H23233).
ACCESSION   CR936687
VERSION     CR936687.1  GI:60219598
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2768)
  AUTHORS   Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2005) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp686H23233) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686H23233
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .2768
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686H23233"
                     /tissue_type="small intestine"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
ORIGIN

  Query Match             10.3%;  Score 205.2;  DB 4;  Length 2768;
  Best Local Similarity   80.6%;  Pred. No. 6.2e-18;
  Matches  253;  Conservative    0;  Mismatches   58;  Indels    3;  Gaps    1;

Qy        372 AATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACAC 431
              |||  |  |||  |  ||||   | |||||||||| || ||||  || |||||||||| |||
Db       2447 AATGAAAAAAAATATTTGCCCTTGGCCAGGCGCGGTGGCTCACGCTTGTAATCCCAGCAC 2506

Qy        432 TTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAA 491
              |||||||||| |||||||| |||||||||||||   ||||||| ||||||||||||||||
Db       2507 TTTGGGAGGCAAAGGTGGGTGGATCACCTGAGGTTGGGAGTTTGAGACCAGCCTGGCCAA 2566

Qy        492 CATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCAC 551
              |||||||||||||  | ||||||||||  ||||||||||  |||||||||| ||||||
Db       2567 CATGGTGAAACCCCATATCTACTAAAAATACAAAAATTAGCGGGGTGTGGTGGTATGCAC 2626

Qy        552 CTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGA 611
              |||||||||| || ||| ||||| |||||  |||||||||||||||||| |||| | |||
Db       2627 CTGTAGTCCCACCTGCTCCGGAGGCTGAGGCTGGAGAATTGCTTGAACCCAGGACGGGGA 2686

Qy        612 GGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCA 668
              ||||||||| | | |||||   | || ||||||||||||||  |  |||  | |||  |
Db       2687 GGTTGCAGTGAGCTGAGATCACGCCATTGCACTCCAGCCTGGGTAACGAGAGAAACTGTA 2746

Qy        669 TAAAGACAACAAAA 682
              |    | || ||||
Db       2747 TTTCAAAAAAAAAA 2760


RESULT 16
AQ037381
LOCUS       AQ037381               455 bp    DNA     linear   GSS 11-JUL-1998
DEFINITION  CIT-HSP-2335L20.TR CIT-HSP Homo sapiens genomic clone 2335L20,
            genomic survey sequence.
ACCESSION   AQ037381
VERSION     AQ037381.1  GI:3303213
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
```

```
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2335L20.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2335L20"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             10.2%;  Score 205;  DB 9;  Length 455;
  Best Local Similarity   80.4%;  Pred. No. 1.3e-17;
  Matches  266;  Conservative    0;  Mismatches   60;  Indels    5;  Gaps    2;

Qy        356 AAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTG 415
              |||| | | | |       | | ||    ||    | || |||||||||||||| |||| ||
Db         98 AAAAATTAAATAAAGTGTCTTTAAACAGACTCACACATAGGCCAGGCGCAGTGGCTCATG 157

Qy        416 CCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTA 475
              ||||||||||||  ||||||||||||| ||| ||| ||||||| |||||||||||||| |
Db        158 CCTGTAATCCCAGTACTTTGGGAGGCCGAGGCGGGTGGATCACTTGAGGTCAGGAGTTCA 217

Qy        476 AGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAG 535
              |||||||||||||||||||| ||| ||||||||||||||||  ||||||||||||||||
Db        218 AGACCAGCCTGGCCAACATGGGGAAGCCCTGTCTCTACTAAAAATACAAAAATTAGCCAG 277

Qy        536 GTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTT 595
              | ||||||||  || |||||| ||||| |||||||||||| ||||| |||||||| ||||
Db        278 GCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTT 337

Qy        596 GAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGG-- 650
              |||||| ||||| |||||||||||| | ||||||   || |||| ||||||||||||
Db        338 GAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGTGCCACTACACTCCAGCCTGGGC 397

Qy        651 CAACAGAGCAAGACTCCATAAAGACAACAAA 681
                |  ||||| | |||||||     | ||||||
Db        398 TACAAGAGCGAAACTCCATCTCAAAAACAAA 428


RESULT 17
CD699526
LOCUS       CD699526                462 bp    mRNA    linear   EST 25-JUN-2003
DEFINITION  EST16162 human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION   CD699526
VERSION     CD699526.1  GI:32229113
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and
            Zeng,Y.-X.
  TITLE     Transcriptional Gene Expression Profile of Human Nasopharynx
  JOURNAL   Unpublished (2003)
COMMENT     Contact: YiXin Zeng
            Cancer Center
            Sun Yat-sen University
            651 DongFeng Road East, GuangZhou 510060, China
            Tel: 86-1380-9770-743
            Fax: 86-20-8775-4506
            Email: yxzeng@gzsums.edu.cn.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="normal nasopharynx"
                     /clone_lib="human nasopharynx"
                     /note="ESTs generated from a normal nasopharynx cDNA
                     library from southern Chinese"
ORIGIN

  Query Match             10.2%;  Score 205;  DB 6;  Length 462;
  Best Local Similarity   84.1%;  Pred. No. 1.3e-17;
  Matches  243;  Conservative    0;  Mismatches   45;  Indels    1;  Gaps    1;

Qy        395 GGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGA 454
              |||| ||||  || |||| |||||| ||||||| |||||||||||||| ||||||||||
Db         34 GGCCGGGCGGGGTGGCTCATGCCTGCAATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGA 93

Qy        455 TCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACT 514
              |||| |||||||||||||| |||| ||||||||||||||||||||||||  ||||||||
Db         94 TCACTTGAGGTCAGGAGTTTGAGACTAGCCTGGCCAACATGGTGAAACCCCATCTCTACT 153

Qy        515 AAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAG 574
              | |||  ||||||||||||||||||||||||  |||||||| ||||| |||||| ||||
Db        154 AGAAATACAAAAATTAGCCAGGTGTGGTGGCACACACCTGTAATCCCAGCTACTCGGGAG 213

Qy        575 GTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCA 634
              | |||||||||||||| | |||| ||  |||||  |||||||||| |   |||| | ||
Db        214 GCTGAGGGAGGAGAATCGTTTGAGCCCTGGAGGCAGAGGTTGCAGTGAGTGGAGACGCCA 273

Qy        635 CTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||||||||||| ||  |||||||||| |||| |    | || ||||
Db        274 CTGCACTCCAGCCTGGGTGACAGAGCAAGGCTCCGTCTCCAAAAAAAAA 322


RESULT 18
BZ612424
LOCUS       BZ612424                693 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHADM75TR Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-22M5, genomic survey sequence.
ACCESSION   BZ612424
VERSION     BZ612424.1  GI:31520985
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
  PUBMED    12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
```

```
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .693
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /clone="MCF7_1-22M5"
                      /sex="female"
                      /clone_lib="Human MCF7 breast cancer cell line library
                      (MCF7_1)"
                      /note="Vector: pECBAC1; Site_1: HindIII; This library was
                      constructed from MCF7 breast cancer cell line by Amplicon
                      Express (http://www.genomex.com) using their standard
                      procedure."
ORIGIN

  Query Match             10.2%;  Score 205;  DB 9;  Length 693;
  Best Local Similarity   81.2%;  Pred. No. 1.1e-17;
  Matches  251;  Conservative    0;  Mismatches   55;  Indels    3;  Gaps    1;

Qy        377 TTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGG 436
              | ||| |    | |    |||  || |  || |||| ||||||||||||||| ||||||||
Db         90 TAAAATAAGCATAATTAAGGCGGGGGGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGG 149

Qy        437 GAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGG 496
              |||||||||||||||||||||||||||||| ||| || ||||||  ||||||||| |||
Db        150 GAGGCCAAGGTGGGCGGATCACCTGAGGTCAAGAGCTTGAGACCATTCTGGCCAACGTGG 209

Qy        497 TGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTA 556
              ||||||||  ||||||||||||   |||||||||| |||  ||||||  ||  ||||||
Db        210 TGAAACCCCTTCTCTACTAAAATTACAAAAATTAGCTAGGCATGGTGGTGTGTGCCTGTA 269

Qy        557 GTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTG 616
              |||  |||||||||||||| ||||| |||||||| |||||||| |||||| ||||||||
Db        270 CTCCTAACTACTCAGGAGGCTGAGGCAGGAGAATGGCTTGAACCCAGGAGGCGGAGGTTG 329

Qy        617 CAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAG 673
              ||||  | ||||||   || |||||||||||||||||| |||||||| |||||||
Db        330 CAGTGAGCCGAGATTGTGCCACTGCACTCCAGCCTGGCGACAGAGCAAAACTCCATCTCA 389

Qy        674 ACAACAAAA 682
              | || ||||
Db        390 AAAAAAAAA 398


RESULT 19
BE787136
LOCUS       BE787136                 704 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601476688F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3879788 5',
            mRNA sequence.
ACCESSION   BE787136
VERSION     BE787136.1  GI:10208334
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 704)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov

            Plate: LLAM9645  row: o  column: 21
            High quality sequence stop: 615.
FEATURES              Location/Qualifiers
     source           1. .704
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3879788"
                      /tissue_type="large cell carcinoma"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="NIH_MGC_68"
                      /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.8 kb.   Library constructed by Life
                      Technologies. "
ORIGIN

  Query Match             10.2%;  Score 205;  DB 2;  Length 704;
  Best Local Similarity   76.5%;  Pred. No. 1.1e-17;
  Matches  264;  Conservative    0;  Mismatches   80;  Indels    1;  Gaps    1;

Qy        339 TATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCC 398
              |||| |||       |     || |||     | |  || |   | |||| | | |  ||||
Db        273 TATTAAATGGCTGCTATTTAAATAAATATGTGACCTTGTCCTAAAGCTATAATACTGGCC 332

Qy        399 AGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCAC 458
              ||| |||| ||||  |||||||||||||  ||||||||||   ||||||| |||||||
Db        333 GGGCACAGTGGCTCACGCCTGTAATCCCAGAACTTTGGGAGGGTGAGGTGGGTGGATCAC 392

Qy        459 CTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAA 518
              |||||||||||||||| |||||||||||||||||||||||||||||  ||||||||||||
Db        393 CTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAA 452

Qy        519 ACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTG 578
              |  ||||||||||  ||  |||||||| || ||| || ||||| ||||||||||| ||
Db        453 ATACAAAAATTAGCTGGGCATGGTGGCAGGCGCCTATAATCCCAGCTACTCAGGAGGCTG 512

Qy        579 AGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGC 638
              ||| |||||||||||||| ||  ||| ||||||||||||| | |    |   ||||||
Db        513 AGGCAGGAGAATTGCTTGACCCCGGGAAGTGGAGGTTGCAGTTAGCAACCACTCCACTGC 572

Qy        639 ACTCCAGCCTG-GCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||||||| || ||||||| ||||||  |    | || ||||
Db        573 ACTCCAGCCTGAGCGACAGAGCGAGACTCTGTCTCAAAAAAAAA 617


RESULT 20
BX484164/c
LOCUS       BX484164                 555 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686J02240_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686J02240 5', mRNA sequence.
ACCESSION   BX484164
VERSION     BX484164.1  GI:31945700
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Bloecker,H., Boecher,M., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     EST (Bloecker,H., Boecher,M., Mewes,H.W., Weil,B., Amid,C., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
```

```
            No s1 sequence available.
            This clone (DKFZp686J02240) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686J02240"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             10.2%;  Score 204.4;  DB 5;  Length 555;
  Best Local Similarity   75.7%;  Pred. No. 1.4e-17;
  Matches  280;  Conservative    0;  Mismatches   86;  Indels    4;  Gaps    2;

Qy        317 TTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTA 376
              | | ||| ||   | || | | || || | |       || |||  |      |  |
Db        413 TATCAGAATTGCATGTTGCTAGTAGTTTACTATCAAAACAACACTTGTGTTTCAAGCTGT 354

Qy        377 TTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGG 436
              | ||| |    | |||| ||||| |    || ||||   ||||||||||||||| |||||
Db        353 TAAAAAATGAATAAAATAGGCCAAGTATGGTGGCTCACACCTGTAATCCCAACATTTTGG 294

Qy        437 GAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGG 496
              |||||| ||| |||||||||| |||||||||||||  | ||||||||||||||||||
Db        293 GAGGCCGAGGCGGGCGGATCACTTGAGGTCAGGAGTTCGACACCAGCCTGGCCAACATGG 234

Qy        497 TGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTA 556
              ||||||||  |||||||||||||  |||||||||||| ||  |||| ||||||  ||||||
Db        233 TGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCCTGGTAGCATGCTCCTGTA 174

Qy        557 GTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTG 616
              ||||||  ||||||||||||   |||| |||||||||||||||  | ||   ||||||  |||
Db        173 GTCCCAGCTACTCAGGAGGCCGAGGCAGGAGAATTGCTTGAACTTGGGCAGTGGAGCTTG 114

Qy        617 CAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAA 672
              ||||  ||| |||   ||| |||||||||||||||| ||||||||||||||||||||
Db        113 CAGTTACCTGAGATCATGCCACTGCACTCCAGCCTGGGCAACAGAGCAAGACTCCATCTC 54

Qy        673 GACAACAAAA 682
              | |  ||||
Db         53 AAAATTAAAA 44


RESULT 21
BZ599806
LOCUS       BZ599806                 829 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHADX62TR Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-24L4, genomic survey sequence.
ACCESSION   BZ599806
VERSION     BZ599806.1  GI:31508268
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 829)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .829
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-24L4"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN

  Query Match             10.2%;  Score 204.4;  DB 9;  Length 829;
  Best Local Similarity   80.3%;  Pred. No. 1.2e-17;
  Matches  265;  Conservative    0;  Mismatches   61;  Indels    4;  Gaps    2;

Qy        358 AACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCC 417
              |||   ||||   |||| ||| | ||  |  |    ||||||| ||||| ||||   ||
Db        354 AACAGGTACAAAAAAATAGTTAGAAAGAATGAATTCAGGCCAGGTGCAGTGGCTCACACC 413

Qy        418 TGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAG 477
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        414 TGTAATCCCAACACTTTGGGAGGCCACGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAG 473

Qy        478 ACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGT 537
              |||||||||||||| |||||||||||| ||||||||||||   |||||||||| ||
Db        474 ACCAGCCTGGCCAATATGGTGAAACCCCGTCTCTACTAAAAATAAAAAAATTAGCCGGGC 533

Qy        538 GTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGA 597
              ||||||  |   |||||| ||||| ||| |  ||||| ||||| ||||||| ||||||
Db        534 ATGGTGGTGTAAGCCTGTAATCCCAGCTATTTGGGAGGCTGAGGCAGGAGAATCGCTTGA 593

Qy        598 ACCTAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCAC---TGCACTCCAGCCT-GGCAA 653
              ||  |||||||||||||||| | | |||||  |||   |||||||||||| |||||
Db        594 ACTCAGGAGGTGGAGGTTGCAGTGAGCTGAGATTGCACCATTGCACTCCAGCCTGGGCAA 653

Qy        654 CAGAGCAAGACTCCATAAAGACAACAAAAG 683
              |||||  ||||||| |    | || |||||
Db        654 CAGAGTGAGACTCCGTCTCAAAAAAAAAG 683


RESULT 22
AI733856/c
LOCUS       AI733856                 458 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  zo19c03.y5 Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:587332 5' similar to contains Alu repetitive element;, mRNA
            sequence.
ACCESSION   AI733856
VERSION     AI733856.1  GI:5054969
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: zo19c03.x5
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read is a RESEQUENCE of a previously sequenced human clone
            Original clone citation: WashU-Merck EST Project
            This read has been verified (found to hit its original self in the
            correct orientation)
            Insert Length: 1908   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 381.
FEATURES              Location/Qualifiers
     source           1. .458
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:4619634"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:587332"
                      /tissue_type="tumor"
                      /cell_line="T84 carcinoma cell line"
                      /lab_host="SOLR cells (kanamycin resistant)"
                      /clone_lib="Stratagene colon (#937204)"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. T-84 colonic epithelial cell line.  Average
                      insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTTT 3'"
ORIGIN

  Query Match             10.2%;  Score 204.2;  DB 1;  Length 458;
  Best Local Similarity   75.3%;  Pred. No. 1.6e-17;
  Matches  281;  Conservative    0;  Mismatches   88;  Indels    4;  Gaps    2;

Qy        314 AGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAA 373
              || | ||    || ||   |  | ||    | |  ||   | |      |    | ||||
Db        398 AGTTATAGACATTGTTGGGTTAACCTTGGGAGTCCAAAGCTCATCCTAAGAGGAATTAAT 339

Qy        374 TTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTT 433
                ||       |||    ||||||||||||  || |||| ||||||||||||||| |||||
Db        338 AATATATCTTTTTTTTTTTTGGGCCAGGCGTGGTGGCTCATGCCTGTAATCCCAGCACTT 279

Qy        434 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACA 493
              |||||| || |||  || ||||||||||||||||||||||  |||||||||| ||||||
Db        278 TGGGAGTCCGAGGCAGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAACA 219

Qy        494 TGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCT 553
              |||||||||||||| ||||||||||  ||||||||||  || ||||||||||||  |||
Db        218 TGGTGAAACCCTGTTTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGCATGCGCCT 159

Qy        554 GTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGG 613
               || ||||| |||||||||| || ||||| |||||||| |||||||||  ||||||||||
Db        158 ATAATCCCAGCTACTCAGGGGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGTGGAGG 99

Qy        614 TTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCAT 669
              ||||||| | ||||||   || ||||||||||||||| ||  |||||||||||||||||
Db         98 TTGCAGTGAGCCGAGATCGTGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCAT 39

Qy        670 AAAGACAACAAAA 682
                | || ||||
Db         38 CTCAAAAAAAAAA 26


RESULT 23
AA808875/c
LOCUS       AA808875                390 bp    mRNA    linear   EST 18-FEB-1998
DEFINITION  nt86h02.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:1205427
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AA808875
VERSION     AA808875.1  GI:2878281
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 929   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 344.
FEATURES              Location/Qualifiers
     source           1. .390
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1205427"
                      /sex="Male"
                      /dev_stage="45 years old"
                      /lab_host="DH10B"
                      /clone_lib="NCI_CGAP_Pr3"
                      /note="Vector: pAMP10; Site_1: Not1; Site_2: EcoRI; 1st
                      strand cDNA was primed with oligo(dT)17 on 50 ng of
                      DNAse-treated, total cellular RNA obtained from
                      5,000-10,000 microdissected cells
                      histologically-determined to be fully malignant prostate
                      cancer cells.  Double-stranded cDNA was ligated to EcoRI
                      adaptors, 5 cycles of PCR applied to the cDNA with an
                      adaptor-specific primer, and the resulting PCR product
                      subcloned into pAMP10 by the UDG-cloning method (Life
                      Technologies).  Average insert size is 600 bp. NOTE: Not
                      directionally cloned.  This library was constructed by
                      David Krizman."
ORIGIN

  Query Match             10.2%;  Score 203.6;  DB 1;  Length 390;
  Best Local Similarity   80.3%;  Pred. No. 2.1e-17;
  Matches  252;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    1;

Qy        372 AATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACAC 431
              ||| |  ||| |  ||||   | |||||||||| ||   ||  || |||||||||| |||
Db        326 AATGAAAAAAAATATTTGCCCTTGGCCAGGCGCGGTGCGTCACGCTTGTAATCCCAGCAC 267

Qy        432 TTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAA 491
              |||||||||| |||||||| ||||||||||||   |||||||| |||||||||||||||||
Db        266 TTTGGGAGGCAAAGGTGGGTGGATCACCTGAGGCTAGGAGTTTGAGACCAGCCTGGCCAA 207

Qy        492 CATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCAC 551
              |||||||||||||  | ||||||||||  ||||||||||  |||||||||| ||||||
Db        206 CATGGTGAAACCCCATATCTACTAAAAATACAAAAATTAGCGGGGTGTGGTGGTATGCAC 147

Qy        552 CTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGA 611
              ||||| ||||| |||||| ||||| |||||  |||||||||||||||||| |||||| |||
Db        146 CTGTAATCCCACCTACTCCGGAGGCTGAGGCTGGAGAATTGCTTGAACCCAGGAGGGGGA 87

Qy        612 GGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCA 668
              |||||||| | | |||||    | || ||||||||||||||||   |  |||  | |||  |
Db         86 GGTTGCAGTGAGCTGAGATCACGCCATTGCACTCCAGCCTGGGTAACGAGAGAAACTGTA 27

Qy        669 TAAAGACAACAAAA 682
              |    | || ||||
Db         26 TTTCAAAAAAAAAA 13


RESULT 24
```

```
CR739570/c
LOCUS       CR739570                438 bp    mRNA    linear   EST 10-DEC-2004
DEFINITION  CR739570 Homo sapiens library (Ebert L) Homo sapiens cDNA clone
            IMAGp998H22178 ; IMAGE:37996 5', mRNA sequence.
ACCESSION   CR739570
VERSION     CR739570.1  GI:51647348
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Ebert,L., Heil,O., Hennig,S., Korn,B., Neubert,P., Partsch,E.,
            Peters,M., Radelof,U. and Schneider,D.
  TITLE     I.M.A.G.E. cDNA Clone Collection
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Email: www.rzpd.de
            RZPD; IMAGp998H22178.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
            Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 100
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998H22178 ; IMAGE:37996"
                     /clone_lib="Homo sapiens library (Ebert L)"
ORIGIN

  Query Match             10.2%;  Score 203.6;  DB 7;  Length 438;
  Best Local Similarity   85.4%;  Pred. No. 2e-17;
  Matches  251;  Conservative    0;  Mismatches   39;  Indels    4;  Gaps    2;

Qy        393 TGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCG 452
              |||||  || || || ||||   ||| |||||||| ||||||||||||||||||||||||
Db        298 TGGGCTGGGTGCTGTGGCTCACACCTATAATCCCAGCACTTTGGGAGGCCAAGGTGGGCG 239

Qy        453 GATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTA 512
              |||||||||||||||||||| |||||||||||||||||||||||||||||| |||||||
Db        238 GATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTA 179

Qy        513 CTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGG 572
              |||||||  |||||||||||||||||||||||  ||||||||||||||  ||||| ||
Db        178 CTAAAAATACAAAAATTAGCCAGGTGTGGTGGTGTGCACCTGTAGTCCCAGGTACTCGGG 119

Qy        573 AGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT-- 630
              ||| ||||| |||||||| |||||||||| ||||||||||||||||||| | || ||||
Db        118 AGGCTGAGGCAGGAGAATAGCTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATCG 59

Qy        631 -GTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
               | ||||||||||||||| ||  ||||||  ||||||  |     | || ||||
Db         58 CGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCTGTCTTAAAAAAAAAA 5


RESULT 25
BF854170
LOCUS       BF854170                545 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  MR2-EN0092-261200-001-e11 EN0092 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF854170
VERSION     BF854170.1  GI:12241914
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  PUBMED    10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-EN0092-
            261200-001-e11&t3=2000-12-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 22
            High quality sequence stop: 532.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="EN0092"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

  Query Match             10.2%;  Score 203.4;  DB 2;  Length 545;
  Best Local Similarity   80.9%;  Pred. No. 1.9e-17;
  Matches  250;  Conservative    0;  Mismatches   56;  Indels    3;  Gaps    1;

Qy        377 TTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGG 436
              ||  |  || ||    | |||||||||||| | ||   ||| |||||||| ||  ||||
Db        206 TTTTAAGGCATTATGTTTGGCCAGGCGCAGTGGTTCACACCTATAATCCCAGCATGTTGG 265

Qy        437 GAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGG 496
              | |||||||||||||| ||||||||||||||||||||||  |||||||||||||||||||
Db        266 GTGGCCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGG 325

Qy        497 TGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTA 556
              ||||||||  ||||||||||||  ||||||||||||||||  |||||||  ||||||||||
Db        326 TGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGCACCTGTA 385

Qy        557 GTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTG 616
              | ||| || ||  ||||| ||||| |||||||| |||||||||| ||||| |||| |||
Db        386 ATTCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAGATTG 445

Qy        617 CAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAG 673
              |||| | || |||   || ||||||||||||||||||  ||||||  |||||||||
Db        446 CAGTGAGCCAAGATTGTGCCACTGCACTCCAGCCTGGTGACAGAGTGAGACTCCATCTCA 505

Qy        674 ACAACAAAA 682
              | || ||||
```

```
Db        506 AAAAAAAAA 514


RESULT 26
CR961104
LOCUS       CR961104                526 bp    DNA     linear   GSS 06-JUN-2005
DEFINITION  Homo sapiens BAC end sequence of RZPDB737G012026D from genomic
            library (orig. Pieter J. de Jong library RPCI-11), genomic survey
            sequence.
ACCESSION   CR961104
VERSION     CR961104.1  GI:66991670
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 526)
  AUTHORS   Schwarz,F., Neubert,P., Schneider,D., Peters,M. and Korn,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2005) RZPD Deutsches Ressourcenzentrum fuer
            Genomforschung GmbH, Im Neuenheimer Feld 515, D-69120 Heidelberg,
            Germany
COMMENT     RZPD; RZPDB737G012026D; RP11-69M1;
            derived from Pieter J. de Jong library RPCI-11;
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPDB737G012026D
            RZPDLIB; (Human Genomic Set - RZPD 1.0) RZPD LIB No.737
            http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=737
            http://www.rzpd.de/products/genomicset/
            Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 100
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available from RZPD;
            Contact RZPD (customer.service@rzpd.de) for further information.
            Clone distribution: http://www.rzpd.de/products/genomicset/
            Seq-primer: T7 (TAA-TAC-GAC-TCA-CTA-TAG-GG)
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .526
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RZPDB737G012026D"
                     /sex="Male"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; RPCI-11 Human Male BAC Library"
ORIGIN

  Query Match            10.1%;  Score 203;  DB 11;  Length 526;
  Best Local Similarity  74.4%;  Pred. No. 2.2e-17;
  Matches  270;  Conservative    0;  Mismatches   90;  Indels    3;  Gaps    1;

Qy        380 AAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAG 439
              |||   ||   |   |||||||||| ||   ||   ||||||||||||||||||||||
Db        106 AAGCATTTAAGACCAGGCCAGGCGCGGTGATTCACACCTGTAATCCCAACACTTTGGGAG 165

Qy        440 GCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGA 499
              ||  ||||||| ||||||||||||||| |||||||||||||||||||||||||||||||
Db        166 GCTGAGGTGGGTGGATCACCTGAGGTCAAGAGTTTAAGACCAGCCTGGCCAACATGGTGA 225

Qy        500 AACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTC 559
              ||| ||||||||||||||  ||||||||||   ||  ||||||||    |||||||||
Db        226 AACTCTGTCTCTACTAAAAATACAAAAATTAGTTGGGCATGGTGGCACATGCCTGTAGTC 285

Qy        560 CCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAG 619
              ||| |||||| ||||| ||||| |||||||| ||||||| |  ||| |  ||||||||||
Db        286 CCAGCTACTCGGGAGGCTGAGGCAGGAGAATAGCTTGAAACCGGGAAGCAGAGGTTGCAG 345

Qy        620 TAACCCGAG---ATGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACA 676
              | | |||||   | | ||||||||||||||||||||||||||  ||||||  |    | |

Db        346 TGAGCCGAGAGCACGCCACTGCACTCCAGCCTGGCAACAGAGTGAGACTCTGTCTCAAAA 405

Qy        677 ACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGAAATTCATCTT 736
              | || |  |   |         ||  ||| ||||||  |  | |  ||| ||   |  |
Db        406 AAAATAAATAACATTGATCAGTGATTGGTTATACCTAGTTGAACTATAGATATGAGTTAT 465

Qy        737 TGT 739
              |||
Db        466 GGT 468


RESULT 27
CR859576/c
LOCUS       CR859576                3165 bp    mRNA    linear   HTC 12-NOV-2004
DEFINITION  Pongo pygmaeus mRNA; cDNA DKFZp468A085 (from clone DKFZp468A085).
ACCESSION   CR859576
VERSION     CR859576.1  GI:55730029
KEYWORDS    HTC.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pongo.
REFERENCE   1  (bases 1 to 3165)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp468A085) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp468A085
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3165
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp468A085"
                     /tissue_type="heart"
                     /clone_lib="468 (synonym: phrt1). Vector pSport1_Sfi; host
                     DH10B; sites SfilA + SfilB"
                     /dev_stage="adult"
                     /note="hypothetical protein (Homo sapiens), differentially
                     spliced"
     gene            1. .3165
                     /gene="DKFZp468A085"
     CDS             34. .687
                     /gene="DKFZp468A085"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH91740.1"
                     /db_xref="GI:55730030"
                     /translation="MKRVNSCVKSDEHVLEELETEGERQLKSLLQHQLDTSVSIEECM
                     SKKESFAPGTMYKPFGKEAAGTMTLSQFQTLHEKDQETASLRELGLNETEILIWKSHV
                     SGEKKTKLRATPEAIQNRLQDIEERISERQRILCLPQRFAKSKQLTRREMEIEKSLFQ
                     GADRHSFLKALYYQAYHKKTSADKYMTSMKRKIKLGTKALLVCSQTAMKKYWRLDNF"
ORIGIN

  Query Match            10.1%;  Score 203;  DB 4;  Length 3165;
  Best Local Similarity  74.6%;  Pred. No. 1.1e-17;
  Matches  296;  Conservative    0;  Mismatches   95;  Indels    6;  Gaps    3;

Qy        292 AACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAAT 351
```

```
Db       3125 AAATAGGATTGTCTATGTAGAAAACTCTAAGATATCCATAATAAAACTTTTAGAATAAAA 3066

Qy        352 TTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAA--ATGGGCCAGGCGCAGTAG 409

Db       3065 GAGTTTAGTCACAGGATACAAATTAATATAGAAAAATCAACTTTAGGCCAGGCACAGTGG 3006

Qy        410 CTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGG 469

Db       3005 CTCATGCCTGTAATCCCACCACTATGGGAGGCTGAAGAGGGAGGATCACCTGAGGTCAGG 2946

Qy        470 AGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATT 529

Db       2945 AGTTTGAGAACAGCCTAGCCAATATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATT 2886

Qy        530 AGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAA 589

Db       2885 AGCCAGGCGTGGTGGCGCATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAA 2826

Qy        590 TTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGC 646

Db       2825 TCGCTTGAACCTAGGTGGCGGAGGTTGCAGTGAGCCGAGATAGTGCCACTGCACTCCAGC 2766

Qy        647 CT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682

Db       2765 CTGGGTGACAAAGCAAGAATCCATCTAAAAAAAAAA 2729


RESULT 28
CD701193
LOCUS       CD701193                492 bp    mRNA    linear   EST 25-JUN-2003
DEFINITION  EST17717 human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION   CD701193
VERSION     CD701193.1  GI:32231823
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and
            Zeng,Y.-X.
  TITLE     Transcriptional Gene Expression Profile of Human Nasopharynx
  JOURNAL   Unpublished (2003)
COMMENT     Contact: YiXin Zeng
            Cancer Center
            Sun Yat-sen University
            651 DongFeng Road East, GuangZhou 510060, China
            Tel: 86-1380-9770-743
            Fax: 86-20-8775-4506
            Email: yxzeng@gzsums.edu.cn.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="normal nasopharynx"
                     /clone_lib="human nasopharynx"
                     /note="ESTs generated from a normal nasopharynx cDNA
                     library from southern Chinese"
ORIGIN

  Query Match             10.1%;  Score 202.8;  DB 6;  Length 492;
  Best Local Similarity   87.2%;  Pred. No. 2.4e-17;
  Matches  246;  Conservative    0;  Mismatches   32;  Indels    4;  Gaps    2;

Qy        392 ATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGC 451

Db        203 AGGGACCAGGCGCAATGGCTCACGCCTGTAATTCCAGCACTTTGGGAGGCTGAGGTGGGC 262

Qy        452 GGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCT 511

Db        263 GGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAGCATGGTGAAACCCTGTCTCT 322

Qy        512 ACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAG 571

Db        323 ACTAAAAATACAAAAATTAGCCAGGAGTGGTGGCAGGCACCTGTAATCCCAGCTACTCAG 382

Qy        572 GAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG--- 628

Db        383 GAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATC 442

Qy        629 ATGTCACTGCACTCCAGCCTGG-CAACAGAGCAAGACTCCAT 669

Db        443 ATGCCACTGCACTCCAGCCTGGCCAGCCGAGCAGGACTCCAT 484


RESULT 29
BU957691/c
LOCUS       BU957691                888 bp    mRNA    linear   EST 21-OCT-2002
DEFINITION  AGENCOURT_10620381 NIH_MGC_107 Homo sapiens cDNA clone
            IMAGE:6731767 5', mRNA sequence.
ACCESSION   BU957691
VERSION     BU957691.1  GI:24187263
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 888)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM3058  row: 1  column: 06
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1. .888
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6731767"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_107"
                     /note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             10.1%;  Score 202.8;  DB 5;  Length 888;
  Best Local Similarity   85.6%;  Pred. No. 2e-17;
  Matches  238;  Conservative    0;  Mismatches   37;  Indels    3;  Gaps    1;

Qy        395 GGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGA 454

Db        278 GGCCAGGCATGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGAGCAGA 219

Qy        455 TCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACT 514

Db        218 TCACCTGAGGTCAGGAGTTTGAGAACAGCCTGGTCAACATGGTGAAACCCCATCTCTACT 159
```

```
Qy        515 AAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAG 574
              |||||  |||||||||||||||| |||||||    || |||||| || || ||| |||||||
Db        158 AAAAATACAAAAATTAGCCAGGCGTGGTGGTGGGCGCCTGTAATCTCAGCTAATCAGGAG 99

Qy        575 GTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATG 631
              | ||||| |||||||| |||||||||| ||||| |||||||||||| | |||||   |||
Db         98 GCTGAGGTAGGAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATCATG 39

Qy        632 TCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCAT 669
              ||||| ||||||||||||  ||||||| ||||||| |
Db         38 CCACTGTACTCCAGCCTGGGGACAGAGCGAGACTCCCT 1


RESULT 30
AQ582332/c
LOCUS       AQ582332                742 bp    DNA     linear   GSS 07-JUN-1999
DEFINITION  RPCI-11-451G7.TV RPCI-11 Homo sapiens genomic clone RPCI-11-451G7,
            genomic survey sequence.
ACCESSION   AQ582332
VERSION     AQ582332.1  GI:5009442
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 742)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .742
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7672950"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-451G7"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             10.1%;  Score 202.6;  DB 9;  Length 742;
  Best Local Similarity   83.3%;  Pred. No. 2.2e-17;
  Matches  244;  Conservative    0;  Mismatches   44;  Indels    5;  Gaps    1;

Qy        390 AAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGG 449
              |||  |||  ||| |||||||||   |||||||||||| ||||||||||||  ||||||
Db        566 AAAAAGGCTGGGCACAGTAGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGG 507

Qy        450 GCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCT 509
              | ||||||||||||||||||||||  ||||||| ||| |||| ||||||||| |||||||
Db        506 GTGGATCACCTGAGGTCAGGAGTTCGAGACCAGTCTGACCAATATGGTGAAAACCTGTCT 447
```

```
Qy        510 CTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTC 569
              ||||||||||  |||||||||||| |||||||||||| ||| ||||||||||||| ||||||
Db        446 CTACTAAAAATACAAAAATTAGCCGGGTGTGGTGGCCTGCGCCTGTAGTCCCAGCTACTC 387

Qy        570 AGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA 629
              |||||| ||||  |||||||||||||||||| |||||| ||| |||||||| | | ||||
Db        386 AGGAGGCTGAGACAGGAGAATTGCTTGAACCCAGGAGGCGGAAGTTGCAGTGAGCTGAGA 327

Qy        630 TGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |     ||||    | || | || |||||||||||||||| || ||||  |||
Db        326 T-----TGCATCACTGCATTGCGACAGAGCAAGACTCCATCAAAACAAACAAA 279


RESULT 31
AG190706
LOCUS       AG190706                470 bp    DNA     linear   GSS 06-MAR-2004
DEFINITION  Pan troglodytes DNA, clone: RP43-066F03.T7, genomic survey
            sequence.
ACCESSION   AG190706
VERSION     AG190706.1  GI:45222882
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
            Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
  TITLE     BAC end sequences of Library RP-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 470)
  AUTHORS   Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
            Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
            Bioscience and Biotechnology (KRIBB), Genome Research Center (GRC);
            52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea
            (E-mail:redstone@mail.kribb.re.kr, URL:http://phs.grc.kribb.re.kr/,
            Tel:82-42-866-7181, Fax:82-42-860-4409)
COMMENT     Clones are derived from the chimpanzee BAC library RP-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: T7
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-066F03.T7"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RP-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match             10.1%;  Score 202.4;  DB 10;  Length 470;
  Best Local Similarity   74.0%;  Pred. No. 2.8e-17;
  Matches  284;  Conservative    0;  Mismatches   96;  Indels    4;  Gaps    2;

Qy        357 AAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGC 416
              |||   | |||   | || || | | |         || |||| ||| |||| |||| |||
Db         87 AAAACCAGACAAAGATATCATAAGAAATGAAAACTATAGGCCGGGCACAGTGGCTCATGC 146

Qy        417 CTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAA 476
              ||||||| ||| |||||||||||||| |||  ||||||||||||||||||||||||| ||
Db        147 CTGTAATTCCAGCACTTTGGGAGGCCGAGGCTGGCGGATCACCTGAGGTCAGGAGTTCAA 206
```

```
Qy        477 GACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGG 536
              ||||||||||||||||||||| |||||||  ||||||||||||  ||||||||||  ||
Db        207 GACCAGCCTGGCCAACATGGAGAAACCCCTTCTCTACTAAAAAAACAAAAATTAGCTGGG 266

Qy        537 TGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTG 596
              ||||  |  ||||||||| ||||| |||||| ||||| ||||  |||||||| |||||
Db        267 CATGGTAGGGTGCACCTGTAATCCCAGCTACTCGGGAGGCTGAGACAGGAGAATCGCTTG 326

Qy        597 AACCTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCA 652
              ||||  |||||| |||| ||||||| | |||||   ||| ||||||||||||||| ||| |
Db        327 AACCCGGGAGGCGGAGTTTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGGCG 386

Qy        653 ACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCT 712
              |||||||||||||||||  | | || ||||         ||   |||| |  | |     |
Db        387 ACAGAGCAAGACTCCATCTAAAAAAAAAAAAGAAATGAAAACTGTAGACCAATATCTCTC 446

Qy        713 ATCTTCATTTAAGAAATTCATCTT 736
              ||   ||     | ||   ||| |
Db        447 ATAAACACAGTTGCAAAAAATCCT 470
```

RESULT 32
AQ769842/c
LOCUS AQ769842 449 bp DNA linear GSS 28-JUL-1999
DEFINITION HS_3170_B2_G07_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3170 Col=14 Row=N, genomic survey sequence.
ACCESSION AQ769842
VERSION AQ769842.1 GI:5647958
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 449)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3170 row: N column: 14
Seq primer: T7
Class: BAC ends
High quality sequence stop: 449.
FEATURES Location/Qualifiers
source 1. .449
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3170 Col=14 Row=N"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
ORIGIN

Query Match 10.1%; Score 202.2; DB 9; Length 449;
Best Local Similarity 83.1%; Pred. No. 3e-17;
Matches 255; Conservative 0; Mismatches 48; Indels 4; Gaps 2;

```
Qy        380 AAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAG 439
              || | |  | | ||||||  |||  ||| |||| |||||||||||||| ||||||||||
Db        445 AAAAACAATCACATGGGCTGGGCCTAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAG 386

Qy        440 GCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGA 499
              |||||| |||| |||||||||||||| |||||||  ||| ||||||||||||||||||||
Db        385 GCCAAGATGGGTGGATCACCTGAGGTGAGGAGTTCGAGAGCAGCCTGGCCAACATGGTGA 326

Qy        500 AACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTC 559
              ||||||||||||||||||  ||||||||||||||||||||||    |||||||| ||
Db        325 AACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGTGTGGTGGTGCACACCTGTAATC 266

Qy        560 CCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAG 619
              ||| |||||| ||||| ||||| |||||||||||||    | ||||||  |||| | |||
Db        265 CCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGTGACCAGGAGGCAGAGGCTACAG 206

Qy        620 TAACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGAC 675
              | | | ||||   || |||||||||||| || ||||||||||||||||||||    |
Db        205 TGAACTGAGAGCTTGCCACTGCACTCCAGTCTGGGCAACAGAGCAAGACTCCATCTCAAA 146

Qy        676 AACAAAA 682
              |||||||
Db        145 AACAAAA 139
```

RESULT 33
AQ822778
LOCUS AQ822778 466 bp DNA linear GSS 26-AUG-1999
DEFINITION HS_2205_B2_A12_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2205 Col=24 Row=B, genomic survey sequence.
ACCESSION AQ822778
VERSION AQ822778.1 GI:5785171
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 466)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2205 row: B column: 24
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 466.
FEATURES Location/Qualifiers
source 1. .466
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=2205 Col=24 Row=B"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
ORIGIN

Query Match 10.1%; Score 202.2; DB 9; Length 466;
Best Local Similarity 80.4%; Pred. No. 3e-17;

```
  Matches  262;  Conservative    0;  Mismatches   59;  Indels    5;  Gaps    2;

Qy        362 AATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTA 421
              || ||||  ||| ||| ||  |          || |||||||| |||| |||| ||||||||
Db         65 AAGACACACAAACTATAAACAAAAAGATTGATAGGCCAGGCACAGTGGCTCATGCCTGTA 124

Qy        422 ATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCA 481
              |||||| |||||||||||||||||  ||| ||||||||||||||||||| |||||||
Db        125 ATCCCAGCACTTTGGGAGGCCAAGGCAGGCTGATCACCTGAGGTCAGGAGTTCAAGACCA 184

Qy        482 GCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGG 541
              ||||||||||||||||||||||||||| ||||||||   ||||||| ||||||  |||
Db        185 GCCTGGCCAACATGGTGAAACCCTGTCTGTACTAAAAGTACAAAAATGAGCCAGCCGTGC 244

Qy        542 TGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCT 601
              ||||   |||||||| ||||| ||||||||||  |||||  | ||||||||||||||||
Db        245 TGGCGGACACCTGTAATCCCAGCTACTCAGGAGACTGAGGCTGAAGAATTGCTTGAACCC 304

Qy        602 AGGAGGTGGAGGTTGCAGTAACCCGAGATG---TCACTGCACTCCAGCCTGG--CAACAG 656
              |||||| |||||||||||| | || |||||    ||||||||||||||||   | ||
Db        305 AGGAGGCGGAGGTTGCAGTGAGCCAAGATGGCACCACTGCACTCCAGCCTGGGTTATAAG 364

Qy        657 AGCAAGACTCCATAAAGACAACAAAA 682
              ||||||||| |||    | ||  |||
Db        365 AGCAAGACTTCATCTCAAAAAANAAA 390


RESULT 34
AI479068/c
LOCUS       AI479068                552 bp    mRNA    linear   EST 09-MAR-1999
DEFINITION  tm30g04.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2158134 3'
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AI479068
VERSION     AI479068.1  GI:4372236
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2158134"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_CLL1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
ORIGIN

  Query Match            10.1%;  Score 202.2;  DB 1;  Length 552;
  Best Local Similarity  85.3%;  Pred. No. 2.8e-17;
  Matches  238;  Conservative    0;  Mismatches   38;  Indels    3;  Gaps    1;

Qy        394 GGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGG 453
              || |||||||| || ||||   |||||||||||||||||||||||||| ||||||| ||
Db        290 GGCCCAGGCGCGGTGGCTCACACCTGTAATCCCAACACTTTGGGAGGCCGAGGTGGGTGG 231

Qy        454 ATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 513
              |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db        230 ATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGGGAAACCCTGTCTCTAC 171

Qy        514 TAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGA 573
              ||||||  |||||||||||||| | ||||||    | |||| ||||| ||||||||||
Db        170 TAAAAATACAAAAATTAGCCAGGCCTAGTGGCACATGCTTGTAATCCCAGCTACTCAGGA 111

Qy        574 GGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT--- 630
              || ||||| |||||||| |||||||| |||||||||||| |||||| | |||||||
Db        110 GGCTGAGGCAGGAGAATGACTTGAACCCAGGAGGTGGAGGCTGCAGTGAGCCGAGATTGC 51

Qy        631 GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCAT 669
              | |||||  |||||||  ||  |||||||| |||||||||
Db         50 GCCACTGGCCTCCAGCTGGGTGACAGAGCGAGACTCCAT 12


RESULT 35
AL712937/c
LOCUS       AL712937                552 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686B0794_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686B0794 5', mRNA sequence.
ACCESSION   AL712937
VERSION     AL712937.1  GI:19696293
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Duesterhoeft, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No s1 sequence available.
            This clone (DKFZp686B0794) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686B0794"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match            10.1%;  Score 202.2;  DB 1;  Length 552;
```

```
  Best Local Similarity   85.6%;  Pred. No. 2.8e-17;
  Matches  249;  Conservative    0;  Mismatches   38;  Indels    4;  Gaps    2;

Qy        396 GCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGAT 455
              ||  |||||||| ||||  |||||||||||||| |||||||||||||| ||||||||||||
Db        429 GCTGGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTAAGGTGGGCGGAT 370

Qy        456 CACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTA 515
              || |||| |||||||| || |||||||||||||||||||||||||||||| | ||||||||
Db        369 CATCTGACGTCAGGAGATTGAGACCAGCCTGGCCAACATGGTGAAACCCCGCCTCTACTA 310

Qy        516 AAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGG 575
              ||||  |||||||||||||||||||||||  |  |||||| ||||| ||||||||||||
Db        309 AAAATACAAAAATTAGCCAGGTGTGGTGGCGGGAGCCTGTAATCCCAGCTACTCAGGAGG 250

Qy        576 TTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGT 632
               ||||| |||||||| |||||||||  ||||||||||||||||||||| | | ||||    ||
Db        249 CTGAGGCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGTTGCAGTGAGCAGAGACTGAGT 190

Qy        633 CACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||||||||||||| ||| ||||||| ||||||| |     | || ||||
Db        189 CACTGCACTCCAGCCTGGGCTACAGAGCGAGACTCCGTCTCAAAAAAAAAA 139


RESULT 36
BU517124/c
LOCUS       BU517124                930 bp    mRNA    linear   EST 12-SEP-2002
DEFINITION  AGENCOURT_10164608 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6514671
            5', mRNA sequence.
ACCESSION   BU517124
VERSION     BU517124.1  GI:22824650
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 930)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14090  row: f  column: 16
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .930
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6514671"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
ORIGIN

  Query Match             10.1%;  Score 202.2;  DB 5;  Length 930;
  Best Local Similarity   82.7%;  Pred. No. 2.3e-17;
  Matches  244;  Conservative    0;  Mismatches   48;  Indels    3;  Gaps    1;

Qy        391 AATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGG 450
              | | |||||||||  || ||||  | ||||||||||| ||||| ||||||| |||||||
Db        328 ATTTGGCCAGGCGTGGTGGCTCACACTTGTAATCCCAGCACTTCGGGAGGCCGAGGTGGG 269

Qy        451 CGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTC 510
              ||||||| |||||||||||||||| ||| |||||||||||||| |||||||||| ||||
Db        268 CGGATCATCTGAGGTCAGGAGTTTGAGATCAGCCTGGCCAACGTGGTGAAACCCCGTCTG 209

Qy        511 TACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCA 570
              |||||||||  |||||||||||||||||||||| || |||||||||| ||||| ||||||
Db        208 TACTAAAAATACAAAAATTAGCCAGGTGTGGTGACACGCACCTGTAATCCCAGCTACTCG 149

Qy        571 GGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG-- 628
              || || |||   |||||||| |||||||||   |||| |||||||||||| | |||||
Db        148 GGGGGCTGACACAGGAGAATCGCTTGAACCCGAGAGGCGGAGGTTGCAGTGAGCCGAGAT 89

Qy        629 -ATGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
               ||| |||||||||||||||||| ||||||  |||||||||    | || || |
Db         88 CATGCCACTGCACTCCAGCCTGGCGACAGAGAGAGACTCCATCCCAAAAAAAATA 34


RESULT 37
BF968880
LOCUS       BF968880                935 bp    mRNA    linear   EST 22-JAN-2001
DEFINITION  602270854F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4358940 5',
            mRNA sequence.
ACCESSION   BF968880
VERSION     BF968880.1  GI:12336106
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9997  row: l  column: 13
            High quality sequence stop: 630.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4358940"
                     /tissue_type="adrenal cortex carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_84"
                     /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
                     primed. Average insert size 1.229 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             10.1%;  Score 202;  DB 2;  Length 935;
  Best Local Similarity   76.5%;  Pred. No. 2.4e-17;
  Matches  274;  Conservative    0;  Mismatches   80;  Indels    4;  Gaps    2;

Qy        347 AAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAG 406
              || | | || ||  | |   |  | || ||| ||||    |  || | |||| | | | |
Db         73 AATACTGTATAATTTCACTTATATGAAGTATAAAAGGTAATCAAATTCGGCCGGACACGG 132

Qy        407 TAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTC 466
              | ||||  |||||| |||||| |||||||||||||  ||||| | ||||||||||||||||
Db        133 TGGCTCACGCCTGTGATCCCAGCACTTTGGGAGGCTGAGGTGAGTGGATCACCTGAGGTC 192
```

```
Qy        467 AGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAA 526
              ||||||| |||||||| |||||||||||||||||||   ||||||| |||||  |||||
Db        193 AGGAGTTCAAGACCAGCTTGGCCAACATGGTGAAACCCCATCTCTACCAAAAATACAAAA 252

Qy        527 ATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGA 586
              |||||||||| ||||||||   |||||||| ||||| ||| |||||| | ||||  ||||
Db        253 ATTAGCCAGGCGTGGTGGCGCACACCTGTAATCCCAGCTATTCAGGAAGCTGAGACAGGA 312

Qy        587 GAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCC 643
              ||||  || |||||| |||||   ||||||||||| | | |||||   | ||||||||||
Db        313 GAATCGCCTGAACCCAGGAGACAGAGGTTGCAGTGAGCTGAGATCACGCCACTGCACTCC 372

Qy        644 AGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAA 700
              ||||| |||||||||||||||||||     | || ||| |  |  || |  |  |||
Db        373 AGCCTGGGCAACAGAGCAAGACTCCATCTTAAAAAAAAAGGAGTCAAATTCATAGAAA 430


RESULT 38
AL712324
LOCUS       AL712324                439 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686E0788_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686E0788 5', mRNA sequence.
ACCESSION   AL712324
VERSION     AL712324.1  GI:19695679
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No s1 sequence available.
            This clone (DKFZp686E0788) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686E0788"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             10.1%;  Score 201.8;  DB 1;  Length 439;
  Best Local Similarity   83.0%;  Pred. No. 3.4e-17;
  Matches  254;  Conservative    0;  Mismatches   48;  Indels    4;  Gaps    2;

Qy        381 AGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGG 440
              | | |  |||||  |||||||| ||| |||||   ||||||||||||||||||||||||
Db        132 AAAACAATGAAACAGGCCAGGCACAGAAGCTCACACCTGTAATCCCAACACTTTGGGAGG 191

Qy        441 CCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAA 500
              ||||||  || || |||| ||||||||||||||| ||||||||||||||||||||||||
Db        192 CCAAGGCAGGTGGCTCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAA 251

Qy        501 ACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCC 560
              |||||||||||||||||||  |||||||||||  ||| ||||||   ||||||||| ||
Db        252 ACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGTATGGTGGTGGGCACCTGTAATCT 311

Qy        561 CAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGT 620
              ||  ||||  ||||| ||||||| |||||||||| |||||||||| ||||||||||||||||
Db        312 CAGTTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGTGGAGGTTGCAGT 371

Qy        621 AACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACA 676
              | || |||    || ||||||||||||||| |||||||||| |||||||  |     | |
Db        372 GAGCCAAGATTGTGCCACTGCACTCCAGCCTGGGCAACAGAGTAAGACTCTGTCTCAATA 431

Qy        677 ACAAAA 682
              | ||||
Db        432 AAAAAA 437


RESULT 39
AG037584
LOCUS       AG037584                662 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-013O02.R, genomic survey sequence.
ACCESSION   AG037584
VERSION     AG037584.1  GI:16564457
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 662)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: M13Rev
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .662
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-013O02.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             10.1%;  Score 201.8;  DB 10;  Length 662;
  Best Local Similarity   76.5%;  Pred. No. 2.9e-17;
  Matches  261;  Conservative    0;  Mismatches   77;  Indels    3;  Gaps    1;

Qy        321 AGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAA 380
              | ||   |||  ||       |  || | |  ||| | || || || ||||| |  || |
Db        321 AAAGACTTTTTATAGGCAATTAGAACAGACATTTCACAATTAGTAGACTTACACAATCAG 380

Qy        381 AGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGG 440
              | |   | ||   ||||||||| | || |||| || |||||||||| ||||||||||
Db        381 AAATGCTAGAGGAGGGCCAGGCACCGTGGCTCATGTCTGTAATCCCAGCACTTTGGGAGG 440
```

```
Qy        441 CCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAA 500
              |||||||||| ||||||| ||||||||||||||  ||||||||||||||||||||||||||
Db        441 CCAAGGTGGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAA 500

Qy        501 ACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCC 560
              |||| |||| |||||||||  ||||||||||| ||  ||||||| ||  |||||| |||
Db        501 ACCCCGTCTTTACTAAAAATACAAAAATTAGCTGGGCTTGGTGGCGTGTGCCTGTAATCC 560

Qy        561 CAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGT 620
              || |||||  ||||| ||||| |||||||||||||||||| ||||| |||||||| |||
Db        561 CAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGGAGGAGGTTGAAGT 620

Qy        621 AACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAG 658
              | |  ||||   | || |||||||||| ||||   || ||
Db        621 GAGCAAAGATCGCGCCATTGCACTCCAGTCTGGGGGCACAG 661
```

RESULT 40
AV729893
LOCUS AV729893 678 bp mRNA linear EST 22-NOV-2004
DEFINITION AV729893 HTE Homo sapiens cDNA clone HTEAIE09 5', mRNA sequence.
ACCESSION AV729893
VERSION AV729893.2 GI:55949949
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 678)
AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA HTE clones
JOURNAL Unpublished (2000)
COMMENT On Oct 17, 2000 this sequence version replaced gi:10839314.
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES Location/Qualifiers
source 1. .678
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HTEAIE09"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="HTE"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Query Match 10.1%; Score 201.8; DB 1; Length 678;
Best Local Similarity 79.3%; Pred. No. 2.9e-17;
Matches 264; Conservative 0; Mismatches 64; Indels 5; Gaps 2;

```
Qy        394 GGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGG 453
              |||||  || ||||| ||||  |||||||||||||| ||||||||||||  |||  || ||
Db        239 GGGCCGGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGTGG 298

Qy        454 ATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 513
              |||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db        299 ATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAGCATGGTGAAACCCTGTCTCTAC 358

Qy        514 TAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGA 573
              |||||   |||| |||| | ||||||||||||||||||||||| ||||| ||||||||||
Db        359 TAAAAATATAAAAGTTAGTCGGGTGTGGTGGCATGCACCTGTAATCCCAGCTACTCAGGA 418

Qy        574 GGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---T 630
              || ||||| |||||||| || || ||| |||||||||||||| ||| | | ||||   |
Db        419 GGCTGAGGCAGGAGAATCGCCTGTACCCAGGAGGTGGAGGTTGAAGTGAGCTGAGATTGT 478

Qy        631 GTCACTGCACTCCAGCC--TGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTG 688
              |  ||||||||||||||   |  | ||||| |||||||||    | ||  | | |
Db        479 GCTACTGCACTCCAGCCAGNGTGACAAGAGCGAGACTCCATCTCAAAAAACANAACAAAA 538

Qy        689 AAATTGTGTAAATGAGTTGTACCTATCTTCATT 721
              || || | ||||| | ||      ||  ||||||
Db        539 AAGTTATATAAATCAATTAACATTAATTTCATT 571
```

RESULT 41
BM458632/c
LOCUS BM458632 1102 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6413734 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5497695 5', mRNA sequence.
ACCESSION BM458632
VERSION BM458632.1 GI:18507672
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1102)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12128 row: d column: 16
High quality sequence stop: 679.
FEATURES Location/Qualifiers
source 1. .1102
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5497695"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_85"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
ORIGIN

Query Match 10.1%; Score 201.8; DB 3; Length 1102;
Best Local Similarity 73.0%; Pred. No. 2.4e-17;
Matches 287; Conservative 0; Mismatches 102; Indels 4; Gaps 2;

```
Qy        326 TATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGC 385
              | | |||  |||    |  || || | ||    |          ||  |   | ||  ||
Db        465 TCTCTATGCCAAGCCCTGTATCAAGTGTTTCTCAGCCTACTTATTTTAAACTAAATTAGT 406

Qy        386 TTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAG 445
              |  | || ||||||||   |||||||  |||||| ||||| ||||||||||||||||||
Db        405 CTGCACATCGGCCAGGCATGGTAGCTCACGCCTGTACTCCCAGCACTTTGGGAGGCCAAG 346

Qy        446 GTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCT 505
```

```
                |  |  |||||||||| |||||||||| |||||||||||||||||||||||||||| |
Db        345 GCAGATGGATCACCTGGGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCTT 286

Qy        506 GTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACT 565
              ||||||||||||||  ||||||||||| ||  ||||||| ||  |||||||||||| ||
Db        285 GTCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTGGCGTGTGCCTGTAGTCCCAGCT 226

Qy        566 ACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCC 625
              |||| ||||| ||||| |||||||| |||||||| ||||| |||||||||||| | ||
Db        225 ACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCC 166

Qy        626 GAGAT---GTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAA 681
              |||||     |||||||||||||||| ||| |||||| |||||| |||     | |  ||
Db        165 GAGATTGGTCCACTGCACTCCAGCCTGGGCGACAGAGAAAGACTTCATCTCAAAACAAAC 106

Qy        682 AGCTTTGAAATTGTGTAAATGAGTTGTACCTAT 714
              | |    ||| || ||  |    |   ||| ||
Db        105 AACAACAAAAATGAGTCCACACATCAAACCCAT 73


RESULT 42
AW408767
LOCUS       AW408767                456 bp    mRNA    linear   EST 16-FEB-2000
DEFINITION  UI-HF-BM0-adw-h-10-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone
            IMAGE:3063402 5', mRNA sequence.
ACCESSION   AW408767
VERSION     AW408767.1  GI:6927824
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M.B. Soares Lab
             cDNA Library Arrayed by: M.B. Soares Lab
             DNA Sequencing by: M.B. Soares Lab
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
             The following repetitive elements were found in this cDNA
            sequence:
             121-410, >ALU
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3063402"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LTI)"
                     /clone_lib="NIH_MGC_38"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (2.5-3.5kb). Directionally cloned. Cells provided by Louis
                     M. Staudt, Ph.D. Library preparation by Maria de Fatima
                     Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
ORIGIN

  Query Match           10.1%;  Score 201.6;  DB 1;  Length 456;
  Best Local Similarity   81.6%;  Pred. No. 3.6e-17;
  Matches  258;  Conservative    0;  Mismatches   54;  Indels    4;  Gaps    2;

Qy        371 AAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACA 430
              | |  |||||||| ||| |   ||||| ||| |||| ||||  |||||||||||| ||
Db         97 ACACCATTAAAGAGTTTTAAGGGGGGCCGGGCACAGTGGCTCACGCCTGTAATCCCAGCA 156

Qy        431 CTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCA 490
              ||||||||||  ||| ||||||||||||||||||||||||||| ||||||||||||||||
Db        157 CTTTGGGAGGCTGAGGCGGGCGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCA 216

Qy        491 ACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCA 550
              || ||| |||||||||||||||||||   |||||||||| |  |||||| || |||
Db        217 ACTTGGCGAAACCCTGTCTCTACTAAAAATAGAAAAATTAGCCGGACGTGGTGACAGGCA 276

Qy        551 CCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGG 610
              |||||| ||||| |||||| |||| || || |||||||| |||  |||| |||||||||
Db        277 CCTGTATTCCCAGCTACTCTGGAGTCTGGGGCAGGAGAATCGCTCAAACCCAGGAGGTGG 336

Qy        611 AGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTC 666
              | |||||||||| | |||||     ||||| |||||||||| ||| ||||||| ||||||
Db        337 AAGTTGCAGTAAGCTGAGATCACACCACTGTACTCCAGCCTGGGCGACAGAGCGAGACTC 396

Qy        667 CATAAAGACAACAAAA 682
              ||    | || ||||
Db        397 TATCTCAAAAAAAAAA 412


RESULT 43
AI791156/c
LOCUS       AI791156                524 bp    mRNA    linear   EST 02-JUL-1999
DEFINITION  ab53e09.x5 Stratagene lung carcinoma 937218 Homo sapiens cDNA clone
            IMAGE:844552 3' similar to contains Alu repetitive element;contains
            element TAR1 repetitive element ;, mRNA sequence.
ACCESSION   AI791156
VERSION     AI791156.1  GI:5338872
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Dental Research,
            Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: ab53e09.y5
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read is a RESEQUENCE of a previously sequenced human clone
            Original clone citation: see original entry for original citation
            information
            This 3' resequenced clone has no previous 3' data to verify this
            new read against
            Seq primer: -40UP from Gibco
            High quality sequence stop: 455.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:844552"
                     /tissue_type="lung carcinoma"
                     /cell_line="NCI-H69"
                     /dev_stage="cell line NCI-H69"
                     /lab_host="SOLR (kanamycin resistant)"
                     /clone_lib="Stratagene lung carcinoma 937218"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Small cell carcinoma cell line NCI-H69.  Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
```

```
ORIGIN

  Query Match             10.1%;  Score 201.6;  DB 1;  Length 524;
  Best Local Similarity   82.4%;  Pred. No. 3.4e-17;
  Matches  244;  Conservative    0;  Mismatches   49;  Indels    3;  Gaps    1;

Qy        377 TTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGG 436
              | ||| |     | |    |||  || |  || |||| ||||||||||||||| ||||||||
Db        302 TAAAATAAGCATAATTAAGGCGGGGGGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGG 243

Qy        437 GAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGG 496
              |||||||||||||||||||||||||||||| ||| || ||||||  ||||||||| |||
Db        242 GAGGCCAAGGTGGGCGGATCACCTGAGGTCAAGAGCTTGAGACCATTCTGGCCAACGTGG 183

Qy        497 TGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTA 556
              ||||||||  ||||||||||||   |||||||||| ||| ||||||  ||  ||||||
Db        182 TGAAACCCCTTCTCTACTAAAATTACAAAAATTAGCTAGGCATGGTGGTGTGTGCCTGTA 123

Qy        557 GTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTG 616
              |||  ||||||||||||||| ||||| |||||||| |||||||| |||||| ||||||||
Db        122 CTCCTAACTACTCAGGAGGCTGAGGCAGGAGAATGGCTTGAACCCAGGAGGCGGAGGTTG 63

Qy        617 CAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCAT 669
              |||| | ||||||    || |||||||||||||||||||| ||||||||| |||||||
Db         62 CAGTGAGCCGAGATTGTGCCACTGCACTCCAGCCTGGCGACAGAGCAAAACTCCAT 7


RESULT 44
BG545785
LOCUS       BG545785                 714 bp    mRNA    linear   EST 04-APR-2001
DEFINITION  602573059F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4701302 5',
            mRNA sequence.
ACCESSION   BG545785
VERSION     BG545785.1  GI:13544450
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1536  row: e  column: 15
            High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4701302"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_77"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."

ORIGIN

  Query Match             10.1%;  Score 201.6;  DB 2;  Length 714;
  Best Local Similarity   82.4%;  Pred. No. 3e-17;
  Matches  244;  Conservative    0;  Mismatches   49;  Indels    3;  Gaps    1;

Qy        390 AAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGG 449
              | | ||||  ||| | || |||| ||||||||||||  ||||||||||||  ||||||
Db        360 AGAGGGGCTGGGCACGGTGGCTCATGCCTGTAATCCCGTCACTTTGGGAGGCTGAGGTGG 419

Qy        450 GCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCT 509
              | |||||||||||||||||||||| |||||||||| ||||||||  |||||||||  |||
Db        420 GTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTAGCCAACATAGTGAAACCCCATCT 479

Qy        510 CTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTC 569
              |||||||||  ||||||||||| ||||||||||    ||||||||  |||| |||  |
Db        480 CTACTAAAAATACAAAAATTAGCCGGGTGTGGTGGTGGACACCTGTAACCCCAGCTATGC 539

Qy        570 AGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA 629
              |||||  ||||| ||||||||    ||||||| |||| ||||||||||||| | | ||||
Db        540 GGGAGGCTGAGGCAGGAGAATCATTTGAACCCAGGAAGTGGAGGTTGCAGTGAGCTGAGA 599

Qy        630 ---TGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
                 || |||||||||||||||||| ||||||||||||||||    | || ||||
Db        600 TTGTGCCACTGCACTCCAGCCTGGCGACAGAGCAAGACTCCATCTCAAAAATAAAA 655


RESULT 45
CR936826
LOCUS       CR936826                 3181 bp    mRNA    linear   HTC 23-FEB-2005
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686K21248 (from clone DKFZp686K21248).
ACCESSION   CR936826
VERSION     CR936826.1  GI:60219737
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3181)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Deutschenbaur,S., Schaipp,A.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2005) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Medigenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project.
            This clone (DKFZp686K21248) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686K21248
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3181
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686K21248"
                     /tissue_type="salivary gland"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
ORIGIN

  Query Match             10.1%;  Score 201.6;  DB 4;  Length 3181;
  Best Local Similarity   85.0%;  Pred. No. 1.8e-17;
  Matches  238;  Conservative    0;  Mismatches   39;  Indels    3;  Gaps    1;
```

```
Qy        374 TTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTT 433
              | |||   | |  || | | | |||| ||||||||| ||||  |||||||||||||| |||||
Db       2853 TCATTGTAAATTTTAGTATTAGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTT 2912

Qy        434 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACA 493
              ||||||||| ||||||| |||||||||| |||||||||| |||||||||||||| |||||
Db       2913 TGGGAGGCCGAGGTGGGTGGATCACCTGGGGTCAGGAGTTCAAGACCAGCCTGGGCAACA 2972

Qy        494 TGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCT 553
              ||||||||||  ||| |||||||||  ||||||||||||||||||||||||  ||||||
Db       2973 TGGTGAAACCCCATCTTTACTAAAAATACAAAAATTAGCCAGGTGTGGTGGCGGGCACCT 3032

Qy        554 GTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGG 613
              ||| ||||| |||||||||||| ||||| |||||||| |||||||||  ||||| |||||
Db       3033 GTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCAGGGAGGCGGAGG 3092

Qy        614 TTGCAGTAACCCGAGATGTCAC---TGCACTCCAGCCTGG 650
              ||||||| | |||||||  |||   ||||||||||||||
Db       3093 TTGCAGTGAGCCGAGATCACACCATTGCACTCCAGCCTGG 3132
```

Search completed: January 22, 2006, 21:08:48
Job time : 5429.54 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 01:00:32 ; Search time 6808.48 Seconds
(without alignments)
16706.187 Million cell updates/sec

Title: US-09-728-552A-3_COPY_39000_41000
Perfect score: 2001
Sequence: 1 tgatccaccagccttggcct.....ctaattcagtgaacacttca 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	80622	8	AF222855	AF222855 Homo sapi
	2	1962.6	98.1	112442	8	AL355340	AL355340 Human DNA
c	3	1962.6	98.1	176432	14	AL391648	AL391648 Homo sapi
	4	1961	98.0	69058	8	AF222856	AF222856 Homo sapi
	5	1961	98.0	80202	8	AF222854	AF222854 Homo sapi
	6	1959.4	97.9	41008	6	AX033912	AX033912 Sequence
	7	1959.4	97.9	80155	8	AF042484	AF042484 Homo sapi
	8	1918	95.9	40917	6	AX033911	AX033911 Sequence
	9	683.6	34.2	71032	14	AC084084	AC084084 Homo sapi
c	10	233.6	11.7	122076	8	AC068051	AC068051 Homo sapi
c	11	224.8	11.2	140842	14	AC144747	AC144747 Pan trogl
c	12	224.8	11.2	173623	14	AC144872	AC144872 Pan trogl
	13	224.4	11.2	170787	8	AC121247	AC121247 Homo sapi
c	14	224.4	11.2	174363	8	AC135506	AC135506 Homo sapi
c	15	224.4	11.2	185146	14	AC104311	AC104311 Homo sapi
c	16	224.2	11.2	141521	8	AC073546	AC073546 Homo sapi
c	17	224.2	11.2	142838	14	AC026465	AC026465 Homo sapi
	18	224.2	11.2	180807	8	AC022445	AC022445 Homo sapi
c	19	223.2	11.2	165467	14	AC159033	AC159033 Pan trogl
	20	222.2	11.1	155533	8	AC079177	AC079177 Homo sapi
c	21	222	11.1	106499	8	AC023271	AC023271 Homo sapi
	22	222	11.1	151367	8	AC025750	AC025750 Homo sapi
	23	221.2	11.1	90289	14	AP003164	AP003164 Homo sapi
	24	221.2	11.1	121162	6	AX039602	AX039602 Sequence
	25	221.2	11.1	171612	8	AL358013	AL358013 Human DNA
c	26	221	11.0	185623	14	AC137058	AC137058 Papio anu
	27	221	11.0	202317	8	AC020603	AC020603 Homo sapi
	28	220.6	11.0	174965	8	AC036222	AC036222 Homo sapi
c	29	220	11.0	154369	8	AC004803	AC004803 Homo sapi
	30	220	11.0	162997	8	AC092035	AC092035 Homo sapi
c	31	220	11.0	179146	8	AC067852	AC067852 Homo sapi
c	32	219.2	11.0	841	10	BV651164	BV651164 S217P6186
	33	219.2	11.0	152966	8	AL359713	AL359713 Human DNA
	34	219.2	11.0	159391	14	AC027113	AC027113 Homo sapi
	35	219.2	11.0	180707	14	AC017038	AC017038 Homo sapi
c	36	219	10.9	158135	8	AC025682	AC025682 Homo sapi
	37	219	10.9	248748	14	AC153308	AC153308 Gorilla g
	38	218.8	10.9	102717	8	AC004079	AC004079 Homo sapi
	39	218.8	10.9	133955	8	AC011380	AC011380 Homo sapi
	40	218.8	10.9	178459	14	AC079931	AC079931 Homo sapi
c	41	218.8	10.9	179006	8	AL136365	AL136365 Human DNA
c	42	218.8	10.9	305000	8	HSXDPA	AL590762 Homo sapi
	43	218.4	10.9	103681	8	HS568C11	AL035661 Human DNA
c	44	218.4	10.9	181174	8	AC020634	AC020634 Homo sapi
c	45	218.4	10.9	187825	14	AC161061	AC161061 Pan trogl
	46	218.4	10.9	191971	8	AC146102	AC146102 Pan trogl
	47	218.2	10.9	102818	14	AP000621	AP000621 Homo sapi
c	48	217.6	10.9	81874	8	HS931E15	AL023575 Human DNA
	49	217.6	10.9	120029	14	HSJ282H10	AL132672 Homo sapi
c	50	217.6	10.9	154442	14	AC083971	AC083971 Homo sapi
	51	217.6	10.9	157493	8	AC044836	AC044836 Homo sapi
	52	217.6	10.9	190045	14	AF295016	AF295016 Homo sapi
c	53	217.6	10.9	201197	14	HS424J12	Z82207 Homo sapien
	54	217.2	10.9	164885	14	AC126227	AC126227 Papio anu
	55	217.2	10.9	170237	8	AC142285	AC142285 Pan trogl
	56	217.2	10.9	193001	14	AC023231	AC023231 Homo sapi
	57	216.6	10.8	135762	14	AC139785	AC139785 Homo sapi
c	58	216.6	10.8	139475	8	AC115992	AC115992 Homo sapi
c	59	216.6	10.8	162750	8	AL355365	AL355365 Human DNA
	60	216.4	10.8	122303	8	AC008681	AC008681 Homo sapi
c	61	216.2	10.8	177180	8	HS473B4	Z83826 Human DNA s
c	62	216.2	10.8	180064	8	AC025265	AC025265 Homo sapi
	63	215.8	10.8	193104	14	AC090339	AC090339 Homo sapi
	64	215.6	10.8	153170	8	HS1103G7	AL034548 Human DNA
c	65	215.6	10.8	159116	14	AC074279	AC074279 Homo sapi
	66	215.4	10.8	141659	8	AL139041	AL139041 Human DNA
c	67	215.4	10.8	146376	8	AC009247	AC009247 Homo sapi
	68	215.4	10.8	158608	8	CNS05TDJ	AL356801 Human chr
c	69	215.4	10.8	173877	8	AC108671	AC108671 Homo sapi
c	70	215.4	10.8	177273	8	AC074194	AC074194 Homo sapi
c	71	215.4	10.8	193803	8	AC022819	AC022819 Homo sapi
c	72	215.2	10.8	184435	14	AC141618	AC141618 Homo sapi
	73	215	10.7	176750	8	AC005537	AC005537 Homo sapi
c	74	214.8	10.7	137176	8	AC005017	AC005017 Homo sapi
	75	214.8	10.7	165384	14	AC055735	AC055735 Homo sapi
c	76	214.8	10.7	170346	8	AC020626	AC020626 Homo sapi
c	77	214.8	10.7	187486	14	AC079840	AC079840 Homo sapi
c	78	214.6	10.7	152574	8	AC011195	AC011195 Homo sapi
c	79	214.6	10.7	153147	8	HS13D10	AL021407 Human DNA
	80	214.6	10.7	157493	14	AC027068	AC027068 Homo sapi
	81	214.6	10.7	168128	14	AC024729	AC024729 Homo sapi
c	82	214.6	10.7	170008	8	AC090160	AC090160 Homo sapi
	83	214.6	10.7	172601	14	AC021455	AC021455 Homo sapi
	84	214.6	10.7	191889	14	AC161015	AC161015 Pan trogl
	85	214.6	10.7	195097	8	AC100875	AC100875 Homo sapi
c	86	214.4	10.7	170273	8	CNS05TCD	AL355094 Human chr
c	87	214.4	10.7	325791	6	AR606191	AR606191 Sequence
c	88	214.4	10.7	325791	6	AX234657	AX234657 Sequence
	89	214.2	10.7	180049	8	AC099558	AC099558 Homo sapi
c	90	214	10.7	99257	8	AL365230	AL365230 Human DNA
c	91	214	10.7	158467	8	CNS07EF5	AL512356 Human chr

c	92	214	10.7	161222	14	AC016037 AC016037 Homo sapi
c	93	213.8	10.7	131563	8	HSE12310 BX322561 Homo sapi
c	94	213.8	10.7	140856	8	AC002430 AC002430 Human BAC
c	95	213.8	10.7	159849	14	AC021454 AC021454 Homo sapi
	96	213.8	10.7	169670	8	AL390962 AL390962 Human DNA
c	97	213.8	10.7	170842	8	AC007384 AC007384 Homo sapi
c	98	213.8	10.7	180445	8	AC019171 AC019171 Homo sapi
c	99	213.8	10.7	223201	8	HS53I10 AL133493 Homo sapi
c	100	213.8	10.7	340000	8	HS21C102 AL163302 Homo sapi

ALIGNMENTS

RESULT 1
AF222855
LOCUS AF222855 80622 bp DNA linear PRI 17-JUL-2000
DEFINITION Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION AF222855
VERSION AF222855.1 GI:9246845
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 80622)
AUTHORS Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M., Irvine,D.V., Saffery,R. and Choo,K.H.
TITLE The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification
JOURNAL Genome Res. 10 (6), 832-838 (2000)
PUBMED 10854414
REFERENCE 2 (bases 1 to 80622)
AUTHORS Barry,A.E.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd., Parkville,, Melbourne, Victoria 3052, Australia
REMARK Genomic sequence from human 10q25.2, clonelib=HC
FEATURES Location/Qualifiers
source 1. .80622
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q25.2"
/clone="HC"
repeat_region complement(3. .176)
/rpt_family="Alu"
repeat_region complement(979. .3564)
/rpt_family="Tigger1"
repeat_region 2967. .3094
/rpt_family="Alu"
repeat_region 3664. .3934
/rpt_family="Alu"
repeat_region complement(4920. .5181)
/rpt_family="Alu"
repeat_region complement(6287. .6579)
/rpt_family="Alu"
repeat_region 6625. .6966
/rpt_family="MLT2B2"
repeat_region complement(9215. .9490)
/rpt_family="Alu"
repeat_region complement(11280. .11565)
/rpt_family="Alu"
repeat_region complement(12428. .12675)
/rpt_family="Alu"
repeat_region complement(14304. .14583)
/rpt_family="Alu"
repeat_region 17503. .17794
/rpt_family="Alu"
repeat_region complement(20250. .20533)
/rpt_family="Alu"
repeat_region complement(23031. .23323)
/rpt_family="Alu"
repeat_region complement(24301. .24479)
/rpt_family="MIR"
repeat_region 24951. .25250
/rpt_family="Alu"
repeat_region complement(25987. .26342)
/rpt_family="THE1"
repeat_region 27097. .27335
/rpt_family="MER2"
repeat_region 27538. .27814
/rpt_family="Alu"
repeat_region 27891. .28161
/rpt_family="Alu"
repeat_region 29063. .29334
/rpt_family="Alu"
repeat_region complement(31025. .31093)
/rpt_family="MLT1"
repeat_region complement(31817. .32352)
/rpt_family="MER1"
repeat_region complement(32489. .32607)
/rpt_family="MLT1"
repeat_region 33039. .33316
/rpt_family="Alu"
repeat_region 33469. .33760
/rpt_family="Alu"
repeat_region complement(34527. .34590)
/rpt_family="MIR"
repeat_region 34689. .34995
/rpt_family="Alu"
repeat_region complement(36349. .36437)
/rpt_family="MIR"
repeat_region complement(38796. .39079)
/rpt_family="Alu"
repeat_region 39441. .39693
/rpt_family="Alu"
repeat_region complement(43176. .43237)
/rpt_family="MIR"
repeat_region complement(45203. .45478)
/rpt_family="Alu"
repeat_region 45892. .46025
/rpt_family="MIR"
repeat_region complement(46618. .46851)
/rpt_family="Alu"
repeat_region 48623. .48939
/rpt_family="Alu"
repeat_region complement(49048. .50698)
/rpt_family="L1"
repeat_region complement(50710. .50972)
/rpt_family="Alu"
repeat_region complement(51031. .51192)
/rpt_family="L1"
repeat_region complement(51450. .51739)
/rpt_family="Alu"
repeat_region complement(53270. .53562)
/rpt_family="Alu"
repeat_region 54261. .54808
/rpt_family="LTR12"
repeat_region 59460. .60054
/rpt_family="LTR12"
repeat_region 60349. .60633
/rpt_family="Alu"
repeat_region complement(60669. .60801)
/rpt_family="MIR"
repeat_region complement(61684. .61852)
/rpt_family="MER5"
repeat_region 63466. .63753
/rpt_family="Alu"
repeat_region 64063. .64228
/rpt_family="MER5"
repeat_region 64446. .64521

```
                     /rpt_family="MIR"
     repeat_region   65794. .65883
                     /rpt_family="MER5"
     repeat_region   66772. .66951
                     /rpt_family="MER20"
     repeat_region   complement(67895. .68416)
                     /rpt_family="MER41"
     repeat_region   complement(73192. .73470)
                     /rpt_family="Alu"
     repeat_region   74086. .74367
                     /rpt_family="Alu"
     repeat_region   75916. .76183
                     /rpt_family="Alu"
     repeat_region   complement(77771. .78059)
                     /rpt_family="Alu"
ORIGIN

  Query Match             100.0%;  Score 2001;  DB 8;  Length 80622;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39027 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 39086

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39087 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 39146

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39147 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 39206

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39207 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 39266

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39267 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 39326

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39327 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39386

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39387 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39446

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39447 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39506

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39507 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39566

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39567 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39626

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39627 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39686

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39687 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39746

Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39747 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39806

Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39807 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39866

Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39867 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39926

Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39927 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 39986

Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39987 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 40046

Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40047 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 40106

Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40107 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 40166

Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40167 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 40226

Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40227 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 40286

Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40287 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 40346

Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40347 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40406

Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40407 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40466

Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40467 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40526

Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40527 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 40586

Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40587 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 40646

Qy       1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40647 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 40706

Qy       1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40707 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40766

Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40767 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40826

Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40827 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40886

Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40887 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 40946

Qy       1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40947 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 41006

Qy       1981 TCTAATTCAGTGAACACTTCA 2001
                |||||||||||||||||||||
Db      41007 TCTAATTCAGTGAACACTTCA 41027


RESULT 2
AL355340
LOCUS       AL355340               112442 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-383C6 on chromosome 10 Contains
            the TRUB1 gene for TruB pseudouridine (psi) synthase homolog 1 (E.
            coli), a novel gene and a CpG island, complete sequence.
ACCESSION   AL355340
VERSION     AL355340.17  GI:21436504
KEYWORDS    HTG; TRUB1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 112442)
  AUTHORS   Leongamornlert,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 17, 2002 this sequence version replaced gi:16416169.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-383C6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .112442
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIB753C06383"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-383C6"
                     /clone_lib="RPCI-11.2"
     misc_feature    2000
                     /note="Clone_right_end: RP11-106M7"
     gene            join(24099. .24445,28551. .28649,37000. .37055,45632. .45713,
                     56274. .56346,58041. .58180,60235. .60291,61029. .63577)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
     mRNA            join(24099. .24445,28551. .28649,37000. .37055,45632. .45713,
                     56274. .56346,58041. .58180,60235. .60291,61029. .63577)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
                     /product="TruB pseudouridine (psi) synthase homolog 1 (E.
                     coli)"
                     /note="match: ESTs: AI208778.1 BQ775335.1
                     match: cDNAs: AF448144.1 BC030601.1"
     CDS             join(24160. .24445,28551. .28649,37000. .37055,45632. .45713,
                     56274. .56346,58041. .58180,60235. .60291,61029. .61285)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
                     /standard_name="OTTHUMP00000020554"
                     /note="match: proteins: Q8WWH5 Q9D331"
                     /codon_start=1
                     /product="TruB pseudouridine (psi) synthase homolog 1 (E.
                     coli)"
                     /protein_id="CAH73219.1"
                     /db_xref="GI:55663184"
                     /db_xref="Genew:16060"
                     /db_xref="GOA:Q8WWH5"
                     /db_xref="InterPro:IPR002501"
                     /db_xref="InterPro:IPR004510"
                     /db_xref="UniProt/TrEMBL:Q8WWH5"
                     /translation="MAASEAAVVSSPSLKTDTSPVLETAGTVAAMAATPSARAAAAVV
                     AAAARTGSEARVSKAALATKLLSLSGVFAVHKPKGPTSAELLNRLKEKLLAEAGMPSP
                     EWTKRKKQTLKIGHGGTLDSAARGVLVVGIGSGTKMLTSMLSGSKRYTAIGELGKATD
                     TLDSTGRVTEEKPYDKITQEDIEGILQKFTGNIMQVPPLYSALKKDGQRLSTLMKRGE
                     VVEAKPARPVTVYSISLQKFQPPFFTLDVECGGGFYIRSLVSDIGKELSSCANVLELT
                     RTKQGPFTLEEHALPEDKWTIDDIAQSLEHCSSLFPAELALKKSKPESNEQVLSCEYI
                     TLNEPKREDDVIKTC"
     gene            join(24292. .24445,28551. .28649,37000. .37055,45632. .45906)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-002"
     mRNA            join(24292. .24445,28551. .28649,37000. .37055,45632. .45906)
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-002"
                     /product="TruB pseudouridine (psi) synthase homolog 1 (E.
                     coli)"
                     /note="match: ESTs: AI418848.1 AI422325.1 AI640751.1
                     BE858241.1"
     polyA_signal    61455. .61460
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
     polyA_site      61479
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
     polyA_signal    63558. .63563
                     /gene="TRUB1"
                     /locus_tag="RP11-383C6.1-001"
     polyA_site      63582
     gene            join(80563. .80691,81349. .81444,82381. .82499)
                     /locus_tag="RP11-383C6.2-001"
     mRNA            join(80563. .80691,81349. .81444,82381. .82499)
                     /locus_tag="RP11-383C6.2-001"
                     /product="putative novel transcript"
                     /note="match: ESTs: AA774571.1"
     misc_feature    110443
                     /note="Clone_left_end: RP11-338L11"
ORIGIN

  Query Match             98.1%;  Score 1962.6;  DB 8;  Length 112442;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 1990;  Conservative    0;  Mismatches    9;  Indels    2;  Gaps    2;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107622 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 107681

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     107682 GTCGAATGTTTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 107741
```

```
Qy       121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
Db    107742 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 107801

Qy       181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
Db    107802 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGC-TTCAAGGAGGAA 107860

Qy       241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
Db    107861 TAAACCTATGAAAAGTGTTGAGGATAGTGTCTGATATGAAATAAGGA-TTCAACAAGTAG 107919

Qy       301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
Db    107920 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 107979

Qy       361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
Db    107980 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 108039

Qy       421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
Db    108040 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 108099

Qy       481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
Db    108100 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 108159

Qy       541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
Db    108160 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 108219

Qy       601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
Db    108220 TAGGAGGTGGAGGTTGCAGTGACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 108279

Qy       661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
Db    108280 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 108339

Qy       721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
Db    108340 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 108399

Qy       781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
Db    108400 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 108459

Qy       841 TAATATTCTTGCTTCTTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
Db    108460 TAATATTCTTGCTTCTTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 108519

Qy       901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
Db    108520 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 108579

Qy       961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
Db    108580 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 108639

Qy      1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
Db    108640 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 108699

Qy      1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
Db    108700 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 108759

Qy      1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
Db    108760 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 108819

Qy      1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
Db    108820 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 108879

Qy      1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
Db    108880 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 108939

Qy      1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
Db    108940 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 108999

Qy      1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
Db    109000 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 109059

Qy      1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
Db    109060 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 109119

Qy      1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
Db    109120 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAGATAAC 109179

Qy      1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
Db    109180 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCAGAATGGAATCCTGTAGAATGTTCAC 109239

Qy      1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
Db    109240 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTAGTAAACTATC 109299

Qy      1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
Db    109300 TTAACCAAGACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 109359

Qy      1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
Db    109360 AATAATAAATCTGTCCAGTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 109419

Qy      1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
Db    109420 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 109479

Qy      1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
Db    109480 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 109539

Qy      1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
Db    109540 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 109599

Qy      1981 TCTAATTCAGTGAACACTTCA 2001
Db    109600 TCTAATTCAGTGAACACTTCA 109620
```

RESULT 3
AL391648/c

LOCUS	AL391648 176432 bp DNA linear HTG 18-JUL-2001
DEFINITION	Homo sapiens chromosome 10 clone RP11-203A7, 9 unordered pieces.
ACCESSION	AL391648
VERSION	AL391648.12 GI:14669309
KEYWORDS	HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1
AUTHORS	Chapman,J.
TITLE	Direct Submission
JOURNAL	Submitted (17-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,

```
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 10, 2001 this sequence version replaced gi:14596357.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA203A7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 174005 bases at least Q40
            Consensus quality: 174806 bases at least Q30
            Consensus quality: 175273 bases at least Q20
            Insert size: 175632; sum-of-contigs
            Insert size: 155595; 14.1% error; agarose-fp
            Quality coverage: 6.85x in Q20 bases; sum-of-contigs Quality
            coverage: 8.13x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    23099: contig of 23099 bp in length
            *    23100    23199: gap of 100 bp
            *    23200    38049: contig of 14850 bp in length
            *    38050    38149: gap of 100 bp
            *    38150    65430: contig of 27281 bp in length
            *    65431    65530: gap of 100 bp
            *    65531   100606: contig of 35076 bp in length
            *   100607   100706: gap of 100 bp
            *   100707   156710: contig of 56004 bp in length
            *   156711   156810: gap of 100 bp
            *   156811   159233: contig of 2423 bp in length
            *   159234   159333: gap of 100 bp
            *   159334   171732: contig of 12399 bp in length
            *   171733   171832: gap of 100 bp
            *   171833   174095: contig of 2263 bp in length
            *   174096   174195: gap of 100 bp
            *   174196   176432: contig of 2237 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176432
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-203A7"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .23099
                     /note="assembly_fragment:01553
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    23200. .38049
                     /note="assembly_fragment:00834
                     fragment_chain:1"
     misc_feature    38150. .65430
                     /note="assembly_fragment:02112
                     fragment_chain:1"
     misc_feature    65531. .100606
                     /note="assembly_fragment:01440
                     fragment_chain:1"
     misc_feature    100707. .156710
                     /note="assembly_fragment:01209
                     fragment_chain:1"
     misc_feature    156811. .159233
                     /note="assembly_fragment:00074"
     misc_feature    159334. .171732
                     /note="assembly_fragment:01455"
     misc_feature    171833. .174095
                     /note="assembly_fragment:01550"
     misc_feature    174196. .176432
                     /note="assembly_fragment:03092"
ORIGIN

  Query Match             98.1%;  Score 1962.6;  DB 14;  Length 176432;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 1990;  Conservative    0;  Mismatches    9;  Indels    2;  Gaps    2;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5866 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 5807

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5806 GTCGAATGTTTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 5747

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5746 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 5687

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       5686 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGC-TTCAAGGAGGAA 5628

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db       5627 TAAACCTATGAAAAGTGTTGAGGATAGTGTCTGATATGAAATAAGGA-TTCAACAAGTAG 5569

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5568 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 5509

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5508 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 5449

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5448 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 5389

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5388 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 5329

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5328 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 5269

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       5268 TAGGAGGTGGAGGTTGCAGTGACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 5209

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5208 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 5149

Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5148 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 5089

Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5088 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 5029

Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5028 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 4969
```

```
Qy       901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4968 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 4909

Qy       961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4908 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 4849

Qy      1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4848 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 4789

Qy      1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4788 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 4729

Qy      1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4728 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 4669

Qy      1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4668 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 4609

Qy      1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4608 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 4549

Qy      1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4548 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 4489

Qy      1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4488 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 4429

Qy      1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4428 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 4369

Qy      1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      4368 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAGATAAC 4309

Qy      1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
             |||||||||||||||||||||||||||||||| |||||||||||||| |||||||||||||
Db      4308 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCAGAATGGAATCCTGTAGAATGTTCAC 4249

Qy      1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
             |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      4248 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTAGTAAACTATC 4189

Qy      1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4188 TTAACCAAGACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 4129

Qy      1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      4128 AATAATAAATCTGTCCAGTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 4069

Qy      1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4068 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 4009

Qy      1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4008 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 3949

Qy      1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3948 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 3889

Qy      1981 TCTAATTCAGTGAACACTTCA 2001
             |||||||||||||||||||||
Db      3888 TCTAATTCAGTGAACACTTCA 3868
```

```
RESULT 4
AF222856
LOCUS       AF222856               69058 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone PnC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222856
VERSION     AF222856.1  GI:9246846
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 69058)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 69058)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Human genomic sequence from 10q25.2, clonelib=PnC
FEATURES             Location/Qualifiers
     source          1. .69058
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="PnC"
     repeat_region   complement(598. .3183)
                     /rpt_family="Tigger1"
     repeat_region   2587. .2715
                     /rpt_family="Alu"
     repeat_region   3283. .3553
                     /rpt_family="Alu"
     repeat_region   complement(4548. .4809)
                     /rpt_family="Alu"
     repeat_region   complement(5919. .6211)
                     /rpt_family="Alu"
     repeat_region   6257. .6599
                     /rpt_family="MLT2B2"
     repeat_region   complement(8848. .9123)
                     /rpt_family="Alu"
     repeat_region   complement(10913. .11198)
                     /rpt_family="Alu"
     repeat_region   complement(12062. .12313)
                     /rpt_family="Alu"
     repeat_region   complement(13957. .14236)
                     /rpt_family="Alu"
     repeat_region   17133. .17424
                     /rpt_family="Alu"
     repeat_region   complement(19882. .20197)
                     /rpt_family="Alu"
     repeat_region   complement(22664. .22956)
                     /rpt_family="Alu"
     repeat_region   complement(23935. .24113)
                     /rpt_family="MIR"
     repeat_region   24585. .24884
                     /rpt_family="Alu"
     repeat_region   complement(25618. .25973)
                     /rpt_family="THE1"
     repeat_region   26732. .26970
                     /rpt_family="MER2"
```

```
     repeat_region   27173. .27449
                     /rpt_family="Alu"
     repeat_region   27526. .27796
                     /rpt_family="Alu"
     repeat_region   28702. .28978
                     /rpt_family="Alu"
     repeat_region   complement(30668. .30736)
                     /rpt_family="MLT1"
     repeat_region   complement(31460. .31993)
                     /rpt_family="MER1"
     repeat_region   complement(32130. .32248)
                     /rpt_family="MLT1"
     repeat_region   32685. .32961
                     /rpt_family="Alu"
     repeat_region   33113. .33403
                     /rpt_family="Alu"
     repeat_region   complement(34169. .34233)
                     /rpt_family="MIR"
     repeat_region   34332. .34639
                     /rpt_family="Alu"
     repeat_region   complement(35993. .36081)
                     /rpt_family="MIR"
     repeat_region   complement(38438. .38721)
                     /rpt_family="Alu"
     repeat_region   39081. .39333
                     /rpt_family="Alu"
     repeat_region   complement(42817. .42877)
                     /rpt_family="MIR"
     repeat_region   complement(44845. .45119)
                     /rpt_family="Alu"
     repeat_region   45532. .45665
                     /rpt_family="MIR"
     repeat_region   complement(46257. .46490)
                     /rpt_family="Alu"
     repeat_region   48262. .48578
                     /rpt_family="Alu"
     repeat_region   complement(48687. .50337)
                     /rpt_family="L1"
     repeat_region   complement(50349. .50611)
                     /rpt_family="Alu"
     repeat_region   complement(50670. .50831)
                     /rpt_family="L1"
     repeat_region   complement(51085. .51374)
                     /rpt_family="Alu"
     repeat_region   complement(52905. .53197)
                     /rpt_family="Alu"
     repeat_region   53896. .54467
                     /rpt_family="LTR12"
     repeat_region   59129. .59723
                     /rpt_family="LTR12"
     repeat_region   60018. .60302
                     /rpt_family="Alu"
     repeat_region   complement(60338. .60470)
                     /rpt_family="MIR"
     repeat_region   complement(61353. .61521)
                     /rpt_family="MER5"
     repeat_region   63135. .63416
                     /rpt_family="Alu"
     repeat_region   63724. .63889
                     /rpt_family="MER5"
     repeat_region   64107. .64182
                     /rpt_family="MIR"
     repeat_region   65459. .65548
                     /rpt_family="MER5"
     repeat_region   66439. .66618
                     /rpt_family="MER20"
     repeat_region   complement(67562. .68084)
                     /rpt_family="MER41"
ORIGIN

  Query Match             98.0%;  Score 1961;  DB 8;  Length 69058;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1989;  Conservative    0;  Mismatches   10;  Indels    2;  Gaps    2;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38669 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 38728

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38729 GTCGAATGTTTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 38788

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38789 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 38848

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      38849 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGC-TTCAAGGAGGAA 38907

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||||||
Db      38908 TAAACCTATGAAAAGTGTTGAGGATAGTGTCTGATATGAAATAAGGA-TTCAACAAGTAG 38966

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38967 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39026

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39027 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39086

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39087 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39146

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39147 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39206

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39207 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39266

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              |||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      39267 TAGGAGCTGGAGGTTGCAGTGACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39326

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39327 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39386

Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39387 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39446

Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39447 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39506

Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39507 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39566

Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39567 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 39626

Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39627 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 39686

Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39687 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 39746
```

```
Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39747 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 39806

Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39807 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 39866

Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39867 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 39926

Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39927 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 39986

Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39987 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40046

Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40047 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40106

Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40107 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40166

Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      40167 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAGATAAC 40226

Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
             ||||||||||||||||||||||||||||||||| ||||||||||||||| |||||||||||
Db      40227 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCAGAATGGAATCCTGTAGAATGTTCAC 40286

Qy       1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      40287 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTAGTAAACTATC 40346

Qy       1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40347 TTAACCAAGACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40406

Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      40407 AATAATAAATCTGTCCAGTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40466

Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40467 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40526

Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40527 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 40586

Qy       1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40587 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 40646

Qy       1981 TCTAATTCAGTGAACACTTCA 2001
             |||||||||||||||||||||
Db      40647 TCTAATTCAGTGAACACTTCA 40667
```

```
RESULT 5
AF222854
LOCUS       AF222854               80202 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222854
VERSION     AF222854.1  GI:9246844
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80202)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 80202)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Human genomic sequence from 10q25.2, clonelib=NC, second release
FEATURES             Location/Qualifiers
     source          1. .80202
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="NC"
     repeat_region   complement(598. .3183)
                     /rpt_family="Tigger1"
     repeat_region   2587. .2715
                     /rpt_family="Alu"
     repeat_region   3283. .3553
                     /rpt_family="Alu"
     repeat_region   complement(4548. .4809)
                     /rpt_family="Alu"
     repeat_region   complement(5919. .6211)
                     /rpt_family="Alu"
     repeat_region   6257. .6599
                     /rpt_family="MLT2B2"
     repeat_region   complement(8848. .9123)
                     /rpt_family="Alu"
     repeat_region   complement(10913. .11198)
                     /rpt_family="Alu"
     repeat_region   complement(12062. .12313)
                     /rpt_family="Alu"
     repeat_region   complement(13957. .14236)
                     /rpt_family="Alu"
     repeat_region   17133. .17424
                     /rpt_family="Alu"
     repeat_region   complement(19882. .20197)
                     /rpt_family="Alu"
     repeat_region   complement(22664. .22956)
                     /rpt_family="Alu"
     repeat_region   complement(23935. .24113)
                     /rpt_family="MIR"
     repeat_region   24585. .24884
                     /rpt_family="Alu"
     repeat_region   complement(25618. .25973)
                     /rpt_family="THE1"
     repeat_region   26732. .26970
                     /rpt_family="MER2"
     repeat_region   27173. .27449
                     /rpt_family="Alu"
     repeat_region   27526. .27796
                     /rpt_family="Alu"
     repeat_region   28702. .28978
                     /rpt_family="Alu"
     repeat_region   complement(30668. .30736)
                     /rpt_family="MLT1"
     repeat_region   complement(31460. .31993)
                     /rpt_family="MER1"
     repeat_region   complement(32130. .32248)
                     /rpt_family="MLT1"
```

```
     repeat_region   32685. .32961
                     /rpt_family="Alu"
     repeat_region   33113. .33403
                     /rpt_family="Alu"
     repeat_region   complement(34169. .34233)
                     /rpt_family="MIR"
     repeat_region   34332. .34639
                     /rpt_family="Alu"
     repeat_region   complement(35993. .36081)
                     /rpt_family="MIR"
     repeat_region   complement(38438. .38721)
                     /rpt_family="Alu"
     repeat_region   39081. .39333
                     /rpt_family="Alu"
     repeat_region   complement(42817. .42877)
                     /rpt_family="MIR"
     repeat_region   complement(44845. .45119)
                     /rpt_family="Alu"
     repeat_region   45532. .45665
                     /rpt_family="MIR"
     repeat_region   complement(46257. .46490)
                     /rpt_family="Alu"
     repeat_region   48262. .48578
                     /rpt_family="Alu"
     repeat_region   complement(48687. .50337)
                     /rpt_family="L1"
     repeat_region   complement(50349. .50611)
                     /rpt_family="Alu"
     repeat_region   complement(50670. .50831)
                     /rpt_family="L1"
     repeat_region   complement(51085. .51374)
                     /rpt_family="Alu"
     repeat_region   complement(52905. .53197)
                     /rpt_family="Alu"
     repeat_region   53896. .54467
                     /rpt_family="LTR12"
     repeat_region   59129. .59719
                     /rpt_family="LTR12"
     repeat_region   60018. .60302
                     /rpt_family="Alu"
     repeat_region   complement(60338. .60470)
                     /rpt_family="MIR"
     repeat_region   complement(61353. .61521)
                     /rpt_family="MER5"
     repeat_region   63135. .63416
                     /rpt_family="Alu"
     repeat_region   63724. .63889
                     /rpt_family="MER5"
     repeat_region   64107. .64182
                     /rpt_family="MIR"
     repeat_region   65459. .65548
                     /rpt_family="MER5"
     repeat_region   66439. .66618
                     /rpt_family="MER20"
     repeat_region   complement(67562. .68084)
                     /rpt_family="MER41"
     repeat_region   complement(72861. .73139)
                     /rpt_family="Alu"
     repeat_region   73756. .74037
                     /rpt_family="Alu"
     repeat_region   75588. .75853
                     /rpt_family="Alu"
     repeat_region   complement(77441. .77729)
                     /rpt_family="Alu"
ORIGIN

  Query Match             98.0%;  Score 1961;  DB 8;  Length 80202;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1989;  Conservative    0;  Mismatches   10;  Indels    2;  Gaps    2;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38669 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 38728

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38729 GTCGAATGTTTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 38788

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38789 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 38848

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      38849 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGC-TTCAAGGAGGAA 38907

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db      38908 TAAACCTATGAAAAGTGTTGAGGATAGTGTCTGATATGAAATAAGGA-TTCAACAAGTAG 38966

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38967 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39026

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39027 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39086

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39087 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39146

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39147 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39206

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39207 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39266

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              |||||| ||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      39267 TAGGAGCTGGAGGTTGCAGTGACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39326

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39327 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39386

Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39387 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39446

Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39447 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39506

Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39507 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39566

Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39567 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 39626

Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39627 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 39686

Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39687 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 39746

Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39747 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 39806
```

```
Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39807 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 39866

Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39867 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 39926

Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39927 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 39986

Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39987 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40046

Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40047 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40106

Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40107 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40166

Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      40167 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAGATAAC 40226

Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
             |||||||||||||||||||||||||||||||||||| |||||||||||||| |||||||||
Db      40227 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCAGAATGGAATCCTGTAGAATGTTCAC 40286

Qy       1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
             |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      40287 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTAGTAAACTATC 40346

Qy       1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40347 TTAACCAAGACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40406

Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      40407 AATAATAAATCTGTCCAGTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40466

Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40467 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40526

Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40527 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 40586

Qy       1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40587 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 40646

Qy       1981 TCTAATTCAGTGAACACTTCA 2001
             |||||||||||||||||||||
Db      40647 TCTAATTCAGTGAACACTTCA 40667


RESULT 6
AX033912
LOCUS       AX033912               41008 bp    DNA     linear   PAT 21-SEP-2000
DEFINITION  Sequence 4 from Patent WO9851790.
ACCESSION   AX033912
VERSION     AX033912.1  GI:10280480
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Cancilla,M.R., Choo,K.H. and Du,S.D.
  TITLE     A novel nucleic acid molecule
  JOURNAL   Patent: WO 9851790-A 4 19-NOV-1998;
            CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
            DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES             Location/Qualifiers
     source          1. .41008
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             97.9%;  Score 1959.4;  DB 6;  Length 41008;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1988;  Conservative    0;  Mismatches   11;  Indels    2;  Gaps    2;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38738 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 38797

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38798 GTCGAATGTTTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 38857

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38858 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 38917

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      38918 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGC-TTCAAGGAGGAA 38976

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |||||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db      38977 TAAACCTATGAAAAGTGTTGAGGATAGTGTCTGATATGAAATAAGGA-TTCAACAAGTAG 39035

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39036 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39095

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39096 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39155

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39156 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39215

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39216 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39275

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39276 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39335

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              |||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      39336 TAGGAGCTGGAGGTTGCAGTGACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39395

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39396 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39455

Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      39456 TTAAGAAATTCATCTTTGTTCATCTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39515

Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39516 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39575

Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      39576 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39635
Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
Db      39636 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAGGAT 39695
Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
Db      39696 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 39755
Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
Db      39756 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 39815
Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
Db      39816 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 39875
Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
Db      39876 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 39935
Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
Db      39936 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 39995
Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
Db      39996 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 40055
Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
Db      40056 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40115
Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
Db      40116 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40175
Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
Db      40176 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40235
Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
Db      40236 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 40295
Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
Db      40296 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCAGAATGGAATCCTGTAGAATGTTCAC 40355
Qy       1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
Db      40356 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTAGTAAACTATC 40415
Qy       1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
Db      40416 TTAACCAAGACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40475
Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
Db      40476 AATAATAAATCTGTCCAGTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40535
Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
Db      40536 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40595
Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
Db      40596 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 40655
Qy       1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
Db      40656 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 40715
```

```
Qy       1981 TCTAATTCAGTGAACACTTCA 2001
Db      40716 TCTAATTCAGTGAACACTTCA 40736
```

```
RESULT 7
AF042484
LOCUS       AF042484               80155 bp    DNA     linear   PRI 22-JUN-2000
DEFINITION  Homo sapiens sequence of the 80 kb core centromere protein binding
            region of the mar del(10) neocentromere corresponding to human
            chromosome 10q25.2, complete sequence.
ACCESSION   AF042484
VERSION     AF042484.1  GI:4205782
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,K.H.
  TITLE     Sequence analysis of an 80 kb human neocentromere
  JOURNAL   Hum. Mol. Genet. 8 (2), 217-227 (1999)
   PUBMED   9931329
REFERENCE   2  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville, Melbourne, Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .80155
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /note="Sequence information obtained from BAC clones
                     prepared from cell line BE2Cl-18-5f"
     repeat_region   complement(join(49. .525,3628. .4545,4849. .5053))
                     /rpt_family="L1MEc"
                     /rpt_type=dispersed
     repeat_region   join(560. .597,3184. .3281)
                     /rpt_family="MER58"
                     /rpt_type=dispersed
     repeat_region   complement(join(598. .2553,2716. .3183))
                     /rpt_family="TIGGER1"
                     /rpt_type=dispersed
     repeat_region   2558. .2581
                     /standard_name="T1"
                     /note="100% homology"
                     /rpt_type=direct
                     /rpt_unit="(caggcacagtgg)2"
     repeat_region   2570. .2715
                     /rpt_family="AluSg/x"
                     /rpt_type=dispersed
     repeat_region   3021. .3050
                     /standard_name="T2"
                     /note="83.3% homology"
                     /rpt_type=direct
                     /rpt_unit="(taacaaagtg)3"
     repeat_region   3283. .3558
                     /rpt_family="AluSx"
                     /rpt_type=dispersed
     repeat_region   4173. .4211
                     /standard_name="T3"
                     /note="82.1% homology"
                     /rpt_type=direct
                     /rpt_unit="(aaaaaaataattt)3"
     repeat_region   complement(4546. .4848)
                     /rpt_family="AluSx"
                     /rpt_type=dispersed
```

```
repeat_region   complement(5909. .6208)
                /rpt_family="AluSg"
                /rpt_type=dispersed
repeat_region   join(6232. .6573,6638. .6688)
                /rpt_family="MLT2CB"
                /rpt_type=dispersed
repeat_region   6598. .6645
                /standard_name="AC24"
                /note="100% homology"
                /rpt_type=tandem
                /rpt_unit="ac"
repeat_region   6598. .6637
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="ca"
repeat_region   7091. .7134
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   7808. .7911
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   8349. .8524
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   complement(8839. .9124)
                /rpt_family="AluJb"
                /rpt_type=dispersed
repeat_region   9180. .9280
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   10070. .10170
                /note="low complexity"
                /rpt_family="purine-rich"
repeat_region   10097. .10136
                /standard_name="GA1"
                /note="82.5% homology"
                /rpt_type=direct
                /rpt_unit="(gaaagaaagg)4"
repeat_region   10175. .10339
                /rpt_family="FRAM"
                /rpt_type=dispersed
repeat_region   complement(10899. .11197)
                /rpt_family="AluSx"
                /rpt_type=dispersed
misc_feature    join(11627. .11757,12434. .12512,13443. .13562)
                /note=">98% identity"
                /db_xref="dbEST:AA774571"
repeat_region   11996. .12028
                /standard_name="AT4"
                /note="100% homology"
                /rpt_type=tandem
                /rpt_unit="(attt)8"
repeat_region   complement(12010. .12312)
                /rpt_family="AluJb"
                /rpt_type=dispersed
repeat_region   12327. .12433
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(13946. .14235)
                /rpt_family="AluSq"
                /rpt_type=dispersed
repeat_region   15130. .15159
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   15164. .15753
                /standard_name="AT28"
                /note="82.1% homology"
                /rpt_type=direct
                /rpt_unit="(atgtatatatgtgtatatagacataaat)21"
repeat_region   join(15170. .15289,15410. .15612)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="ta"
repeat_region   15617. .15731
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="cata"
repeat_region   16004. .16100
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16128. .16364)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16407. .16592)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16657. .16889)
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   16914. .16936
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   17102. .17386
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   17406. .17564
                /standard_name="GA2"
                /note="81.8% homology"
                /rpt_type=direct
repeat_region   17416. .17543
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   17647. .17702
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   complement(19863. .20161)
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   complement(20842. .20913)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(21366. .21414)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   21375. .21639
                /standard_name="CT1"
                /note="75% homology"
                /rpt_type=dispersed
repeat_region   complement(21415. .21536)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="tggaa"
repeat_region   complement(21524. .21646)
                /note="low complexity"
                /rpt_family="purine-rich"
repeat_region   22409. .22646
                /rpt_family="L1M4"
                /rpt_type=dispersed
repeat_region   22560. .22604
                /standard_name="T4"
                /note="80% homology"
                /rpt_type=direct
                /rpt_unit="(aatattacaataatt)3"
```

Query Match 97.9%; Score 1959.4; DB 8; Length 80155;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1988; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38647 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 38706

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      38707 GTCGAATGTTTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 38766
Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
Db      38767 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 38826
Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
Db      38827 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGC-TTCAAGGAGGAA 38885
Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
Db      38886 TAAACCTATGAAAAGTGTTGAGGATAGTGTCTGATATGAAATAAGGA-TTCAACAAGTAG 38944
Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
Db      38945 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39004
Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
Db      39005 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39064
Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
Db      39065 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39124
Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
Db      39125 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39184
Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
Db      39185 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39244
Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
Db      39245 TAGGAGCTGGAGGTTGCAGTGACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39304
Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
Db      39305 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39364
Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
Db      39365 TTAAGAAATTCATCTTTGTTCATCTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39424
Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
Db      39425 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39484
Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
Db      39485 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39544
Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
Db      39545 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAGGAT 39604
Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
Db      39605 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 39664
Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
Db      39665 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 39724
Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
Db      39725 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 39784
Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
Db      39785 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 39844

Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
Db      39845 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 39904
Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
Db      39905 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 39964
Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
Db      39965 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40024
Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
Db      40025 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40084
Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
Db      40085 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40144
Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
Db      40145 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 40204
Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
Db      40205 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCAGAATGGAATCCTGTAGAATGTTCAC 40264
Qy       1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
Db      40265 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTAGTAAACTATC 40324
Qy       1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
Db      40325 TTAACCAAGACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40384
Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
Db      40385 AATAATAAATCTGTCCAGTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40444
Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
Db      40445 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40504
Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
Db      40505 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 40564
Qy       1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
Db      40565 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 40624
Qy       1981 TCTAATTCAGTGAACACTTCA 2001
Db      40625 TCTAATTCAGTGAACACTTCA 40645
```

```
RESULT 8
AX033911
LOCUS       AX033911               40917 bp    DNA     linear   PAT 21-SEP-2000
DEFINITION  Sequence 3 from Patent WO9851790.
ACCESSION   AX033911
VERSION     AX033911.1  GI:10280479
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Cancilla,M.R., Choo,K.H. and Du,S.D.
  TITLE     A novel nucleic acid molecule
  JOURNAL   Patent: WO 9851790-A 3 19-NOV-1998;
            CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
```

```
           DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES             Location/Qualifiers
     source          1. .40917
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             95.9%;  Score 1918;  DB 6;  Length 40917;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1918;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39000 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 39059

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39060 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 39119

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39120 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 39179

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39180 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 39239

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39240 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 39299

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39300 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39359

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39360 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39419

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39420 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39479

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39480 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39539

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39540 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39599

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39600 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39659

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39660 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39719

Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39720 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39779

Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39780 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39839

Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39840 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39899

Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39900 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 39959

Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39960 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 40019

Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40020 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 40079

Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40080 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 40139

Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40140 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 40199

Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40200 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 40259

Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40260 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 40319

Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40320 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40379

Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40380 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40439

Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40440 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40499

Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40500 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 40559

Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40560 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 40619

Qy       1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40620 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 40679

Qy       1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40680 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40739

Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40740 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40799

Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40800 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40859

Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCA 1918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40860 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCA 40917

RESULT 9
AC084084
LOCUS       AC084084               71032 bp    DNA     linear   HTG 25-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-314C19 map 8, LOW-PASS
            SEQUENCE SAMPLING.
```

```
ACCESSION   AC084084
VERSION     AC084084.2  GI:13446278
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 71032)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-314C19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71032)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 25, 2001 this sequence version replaced gi:10799449.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11327
                Center clone name: 314_C_19
            --------------.
            * NOTE: This record contains 85 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      720: contig of 720 bp in length
            *      721      820: gap of 100 bp
            *      821     1557: contig of 737 bp in length
            *     1558     1657: gap of 100 bp
            *     1658     2404: contig of 747 bp in length
            *     2405     2504: gap of 100 bp
            *     2505     3299: contig of 795 bp in length
            *     3300     3399: gap of 100 bp
            *     3400     4153: contig of 754 bp in length
            *     4154     4253: gap of 100 bp
            *     4254     4978: contig of 725 bp in length
            *     4979     5078: gap of 100 bp
            *     5079     5815: contig of 737 bp in length
            *     5816     5915: gap of 100 bp
            *     5916     6648: contig of 733 bp in length
            *     6649     6748: gap of 100 bp
            *     6749     7520: contig of 772 bp in length
            *     7521     7620: gap of 100 bp
            *     7621     8347: contig of 727 bp in length
            *     8348     8447: gap of 100 bp
            *     8448     9227: contig of 780 bp in length
            *     9228     9327: gap of 100 bp
            *     9328    10052: contig of 725 bp in length
            *    10053    10152: gap of 100 bp
            *    10153    10869: contig of 717 bp in length
            *    10870    10969: gap of 100 bp
            *    10970    11718: contig of 749 bp in length
            *    11719    11818: gap of 100 bp
            *    11819    12557: contig of 739 bp in length
            *    12558    12657: gap of 100 bp
            *    12658    13412: contig of 755 bp in length
            *    13413    13512: gap of 100 bp
            *    13513    14255: contig of 743 bp in length
            *    14256    14355: gap of 100 bp
            *    14356    15086: contig of 731 bp in length
            *    15087    15186: gap of 100 bp
            *    15187    15901: contig of 715 bp in length
            *    15902    16001: gap of 100 bp
            *    16002    16711: contig of 710 bp in length
            *    16712    16811: gap of 100 bp
            *    16812    17539: contig of 728 bp in length
            *    17540    17639: gap of 100 bp
            *    17640    18351: contig of 712 bp in length
            *    18352    18451: gap of 100 bp
            *    18452    19162: contig of 711 bp in length
            *    19163    19262: gap of 100 bp
            *    19263    20054: contig of 792 bp in length
            *    20055    20154: gap of 100 bp
            *    20155    20891: contig of 737 bp in length
            *    20892    20991: gap of 100 bp
            *    20992    21720: contig of 729 bp in length
            *    21721    21820: gap of 100 bp
            *    21821    22553: contig of 733 bp in length
            *    22554    22653: gap of 100 bp
            *    22654    23420: contig of 767 bp in length
            *    23421    23520: gap of 100 bp
            *    23521    24234: contig of 714 bp in length
            *    24235    24334: gap of 100 bp
            *    24335    25069: contig of 735 bp in length
            *    25070    25169: gap of 100 bp
            *    25170    25915: contig of 746 bp in length
            *    25916    26015: gap of 100 bp
            *    26016    26751: contig of 736 bp in length
            *    26752    26851: gap of 100 bp
            *    26852    27574: contig of 723 bp in length
            *    27575    27674: gap of 100 bp
            *    27675    28437: contig of 763 bp in length
            *    28438    28537: gap of 100 bp
            *    28538    29258: contig of 721 bp in length
            *    29259    29358: gap of 100 bp
            *    29359    30068: contig of 710 bp in length
            *    30069    30168: gap of 100 bp
            *    30169    30918: contig of 750 bp in length
            *    30919    31018: gap of 100 bp
            *    31019    31770: contig of 752 bp in length
            *    31771    31870: gap of 100 bp
            *    31871    32658: contig of 788 bp in length
            *    32659    32758: gap of 100 bp
            *    32759    33501: contig of 743 bp in length
            *    33502    33601: gap of 100 bp
            *    33602    34347: contig of 746 bp in length
            *    34348    34447: gap of 100 bp
            *    34448    35165: contig of 718 bp in length
            *    35166    35265: gap of 100 bp
            *    35266    35998: contig of 733 bp in length
            *    35999    36098: gap of 100 bp
```

```
            *    36099    36802: contig of 704 bp in length
            *    36803    36902: gap of 100 bp
            *    36903    37632: contig of 730 bp in length
            *    37633    37732: gap of 100 bp
            *    37733    38454: contig of 722 bp in length
            *    38455    38554: gap of 100 bp
            *    38555    39275: contig of 721 bp in length
            *    39276    39375: gap of 100 bp
            *    39376    40113: contig of 738 bp in length
            *    40114    40213: gap of 100 bp
            *    40214    40948: contig of 735 bp in length
            *    40949    41048: gap of 100 bp
            *    41049    41792: contig of 744 bp in length
            *    41793    41892: gap of 100 bp
            *    41893    42630: contig of 738 bp in length
            *    42631    42730: gap of 100 bp
            *    42731    43459: contig of 729 bp in length
            *    43460    43559: gap of 100 bp
            *    43560    44277: contig of 718 bp in length
            *    44278    44377: gap of 100 bp
            *    44378    45120: contig of 743 bp in length
            *    45121    45220: gap of 100 bp
            *    45221    45942: contig of 722 bp in length
            *    45943    46042: gap of 100 bp
            *    46043    46775: contig of 733 bp in length
            *    46776    46875: gap of 100 bp
            *    46876    47609: contig of 734 bp in length
            *    47610    47709: gap of 100 bp
            *    47710    48446: contig of 737 bp in length
            *    48447    48546: gap of 100 bp
            *    48547    49285: contig of 739 bp in length
            *    49286    49385: gap of 100 bp
            *    49386    50178: contig of 793 bp in length
            *    50179    50278: gap of 100 bp
            *    50279    51008: contig of 730 bp in length
            *    51009    51108: gap of 100 bp
            *    51109    51846: contig of 738 bp in length
            *    51847    51946: gap of 100 bp
            *    51947    52672: contig of 726 bp in length
            *    52673    52772: gap of 100 bp
            *    52773    53505: contig of 733 bp in length
            *    53506    53605: gap of 100 bp
            *    53606    54330: contig of 725 bp in length
            *    54331    54430: gap of 100 bp
            *    54431    55166: contig of 736 bp in length
            *    55167    55266: gap of 100 bp
            *    55267    55979: contig of 713 bp in length
            *    55980    56079: gap of 100 bp
            *    56080    56823: contig of 744 bp in length
            *    56824    56923: gap of 100 bp
            *    56924    57626: contig of 703 bp in length

  Query Match             34.2%;  Score 683.6;  DB 14;  Length 71032;
  Best Local Similarity   99.0%;  Pred. No. 1.1e-131;
  Matches  708;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    2;

Qy        860 TTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAATACACTGACTTTACACAAAA 919
              || | ||||||||||||||| ||| ||||||||||||||||||||||||||||||||||
Db      12699 TTTACGAATTGTCTCCCTAG-AACTGTTTGTACCAAACAATACACTGACTTTACACAAAA 12757

Qy        920 TCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGATTTTCATTTGAGAAGGGGCC 979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12758 TCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGATTTTCATTTGAGAAGGGGCC 12817

Qy        980 CATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGTTCTGTTTTTCTGGCTTCTG 1039
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12818 CATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGTTCTGTTTTTCTGGCTTCTG 12877

Qy       1040 GGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTTTCTTTATCCTTCATTTCAA 1099
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12878 GGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTTTCTTTATCCTTCATTTCAA 12937

Qy       1100 TACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACATCAAAAGAAAGAGAAACAGT 1159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12938 TACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACATCAAAAGAAAGAGAAACAGT 12997

Qy       1160 TTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCATTTTGTTCCGCTCTAATTA 1219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12998 TTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCATTTTGTTCCGCTCTAATTA 13057

Qy       1220 GCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGCGTGGTAACACAGGTCTTCA 1279
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      13058 GCCAGTGCATGACTTANAGAGCAGGCAGATGCTTTGAAGGCGTGGTAACACAGGTCTTCA 13117

Qy       1280 TTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGCTGAAGAAAATGGCTCAGGA 1339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13118 TTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGCTGAAGAAAATGGCTCAGGA 13177

Qy       1340 AGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGGTGTTATAAAGCACAAGTCA 1399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13178 AGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGGTGTTATAAAGCACAAGTCA 13237

Qy       1400 AGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTTTCTACGTCAGAAGAGCCAA 1459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13238 AGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTTTCTACGTCAGAAGAGCCAA 13297

Qy       1460 GCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTATTCTAAAGTCATGAACAAG 1519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13298 GCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTATTCTAAAGTCATGAACAAG 13357

Qy       1520 CCTAATTATAAAG-ATAGTTGCTGTGAAGGTGCTGAAATAACTCGATTTTACCAA 1573
              ||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db      13358 CCTAATTATAAAGAATAGTTGCTGTGAAGGTGCTGAGATAACTCGATTTTACCAA 13412


RESULT 10
AC068051/c
LOCUS       AC068051               122076 bp    DNA     linear   PRI 16-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-172J1 from 2, complete sequence.
ACCESSION   AC068051
VERSION     AC068051.7  GI:16799049
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 122076)
  AUTHORS   Cotton,M., Maupin,R., Elliott,G. and Grewal,N.
  TITLE     The sequence of Homo sapiens BAC clone RP11-172J1
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 122076)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 122076)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 122076)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 122076)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 122076)
  AUTHORS   Waterston,R.H.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   7  (bases 1 to 122076)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   8  (bases 1 to 122076)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 8, 2001 this sequence version replaced gi:15638775.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0172J01
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP11-223M6, 200 bp overlap.
            Actual start of this clone is at base position 1 of RP11-172J1;
            actual end is at base position 33760 of RP11-223M6.
FEATURES             Location/Qualifiers
     source          1. .122076
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-172J1"
                     /clone_lib="RPCI-11"
     misc_feature    112630. .112861
                     /note="CpG_island (%GC=60.8, o/e=1.00, #CpGs=21)"
ORIGIN

  Query Match             11.7%;  Score 233.6;  DB 8;  Length 122076;
  Best Local Similarity   61.9%;  Pred. No. 1.8e-38;
  Matches  422;  Conservative    0;  Mismatches  254;  Indels    6;  Gaps    3;

Qy        274 ATATGAAATAAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTATTTATT 333
              | |  ||||||    |  ||  |  |  |  |  ||||  | |  ||    | |  ||||

Db      22195 AAACAAAATAACTTATAAAATTATAAAAATATTATATTATAAAAATATTTTTAAAATATT 22136

Qy        334 ACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAAT 393
              | || ||| | ||||||| ||| ||| || || |      ||||| |||     |||  |

Db      22135 ATAATTATATTATAAAATATTATAAAATATTATAAAAGTATTATAAAAATAATTTTTA-- 22078

Qy        394 GGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGG 453
              |||||||| |  |||||||   ||||||||||| |||||||||||||||||||||| |

Db      22077 GGGCCAGGAGTGGTAGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAG 22018

Qy        454 ATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 513
              |||||||||||||||||||  || ||| |||||||||||| ||||||||  |||||||

Db      22017 ATCACCTGAGGTCAGGAGTTCGAGTCCACCCTGGCCAACATGATGAAACCCCATCTCTAC 21958

Qy        514 TAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGA 573
              ||||||   |||||||| |  |||||||||||  |||||||| | ||| ||||||||||

Db      21957 TAAAAATATAAAAATTAACTGGGTGTGGTGGCACACACCTGTAATTCCAGCTACTCAGGA 21898

Qy        574 GGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG---AT 630
              || ||||| |||||||| ||||||||| |||||||||||||||| | | | |||   ||

Db      21897 GGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGTTGCCGCGAGCTGAGATCAT 21838

Qy        631 GTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGA 689
              | |||||||||||||||| |||||||||||| ||||||||    | || ||||   ||||

Db      21837 GCCACTGCACTCCAGCCTGGGCAACAGAGCAGGACTCCATCTCAAAAAAAAAAAAATTGA 21778

Qy        690 AATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGAAATTCATCTTTGTTCATTTATTT 749
              ||||  ||  |    |   ||  | |          | | | | | ||   | |||

Db      21777 CTTTGTATATCTTTTCTAATCCCTTTATTTGAATTTTCAGTGTTTTATTTTAGGTGTTTT 21718

Qy        750 TTACTTGACATGAGAGCTTCCAGCAATTTTTAATTAAGCCCTCACAGATTTTATGTCACT 809
              |    | | |||     ||  |  |  | ||| |||| |  | |     |  | | |

Db      21717 TCTTATAATATGCTGAATTGAAAAACATCTTATTTAATCTTTTATTAGGTAATTTTAAAC 21658

Qy        810 GGCTATGTGATAAACAAATTATTTGCTAAAATAATATTCTTGCTTCTTTTTTAAGGAATT 869
              |||  |         | |  |     |  | || ||| |   |||| ||   | |   |

Db      21657 CTTTATATTTGTTGTTACTACTGGTTAATTAAAACATTTTACTTTCTATTACCATGTTAT 21598

Qy        870 GTCTCCCTAGAAACGGTTTGTACCAAACAATACACTGACTTTACACAAAATCAGATCTGA 929
              |   | ||    |   |  | ||     |||   ||  || |       ||  |  |

Db      21597 TTTCTCTTACTCTCCCCTACTCCCCTGTTATAACTTGGATTGATCTTTTTTCTCAATTTT 21538

Qy        930 TTGGCAACAGTTGCAGATGTTT 951
              ||  |  | | |  ||| ||||

Db      21537 TTTCCTCCTGGTTTAGAAGTTT 21516


RESULT 11
AC144747/c
LOCUS       AC144747               140842 bp    DNA     linear   HTG 07-JUN-2003
DEFINITION  Pan troglodytes clone CH251-160D7, WORKING DRAFT SEQUENCE, 5
            ordered pieces.
ACCESSION   AC144747
VERSION     AC144747.2  GI:31455634
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 140842)
  AUTHORS   Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
            Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
            Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
            Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
            Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 140842)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 140842)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 7, 2003 this sequence version replaced gi:30725904.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: esb
                 Center clone name: 160D07

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 140151 bases at least Q40
                 Consensus quality: 140382 bases at least Q30
                 Consensus quality: 140426 bases at least Q20
                 Insert size: 132000; agarose-fp
                 Insert size: 140442; sum-of-contigs
                 Quality coverage: 11.84x in Q20 bases; agarose-fp
                 Quality coverage: 11.13x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    60845: contig of 60845 bp in length
            *    60846    60945: gap of unknown length
            *    60946    99599: contig of 38654 bp in length
            *    99600    99699: gap of unknown length
            *    99700   102257: contig of 2558 bp in length
            *   102258   102357: gap of unknown length
            *   102358   128790: contig of 26433 bp in length
            *   128791   128890: gap of unknown length
            *   128891   140842: contig of 11952 bp in length.
FEATURES             Location/Qualifiers
     source          1. .140842
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="CH251-160D7"
                     /clone_lib="CH251"
     misc_feature    1. .101605
                     /note="clone overlaps with GenBank Accession Number
                     AC144872 clone CH251-495I14 (center project name esa)"
     misc_feature    1. .60845
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     gap             60846. .60945
                     /estimated_length=unknown
     misc_feature    60946. .99599
                     /note="assembly_fragment"
     misc_feature    84997. .140842
                     /note="clone overlaps with GenBank Accession Number
                     AC096878 clone RP43-171C12 (center project name cls)"
     gap             99600. .99699
                     /estimated_length=unknown
     misc_feature    99700. .102257
                     /note="assembly_fragment"
     gap             102258. .102357
                     /estimated_length=unknown
     misc_feature    102358. .128790
                     /note="assembly_fragment"
     gap             128791. .128890
                     /estimated_length=unknown
     misc_feature    128891. .140842
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             11.2%;  Score 224.8;  DB 14;  Length 140842;
  Best Local Similarity   73.3%;  Pred. No. 1.1e-36;
  Matches  302;  Conservative    0;  Mismatches  107;  Indels    3;  Gaps    1;

Qy        377 TTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGG 436
              ||||| ||    ||||| |||||||| |||| ||||  |||||||||||| ||||||||
Db       4056 TTAAAAAGGAGAGAAATTGGCCAGGCTCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3997

Qy        437 GAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGG 496
              |||||| ||||||| |||||||||||||||||||||||  ||||||||||||||||||||
Db       3996 GAGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGG 3937

Qy        497 TGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTA 556
              ||||||||||||||||||||||||    |||||||||||||||||||||||| || ||||||
Db       3936 TGAAACCCTGTCTCTACTAAAAATAAAAAAATTAGCCAGGTGTGGTGGCAGGCGCCTGTA 3877

Qy        557 GTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTG 616
              |||||  |||||| ||||| ||||| ||||||||  ||||||||  ||||| ||| ||||
Db       3876 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCCGGGAGGCGGAAGTTG 3817

Qy        617 CAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAG 673
              ||||  | |||||||   |||||||||||||||||||  |||||  ||  ||||||||
Db       3816 CAGTGAGCCGAGATCGCGTCACTGCACTCCAGCCTGGTGACAGAACACGACTCCATCTCA 3757

Qy        674 ACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGAAATTCAT 733
              | ||  |||        || | |    |||   |       ||   | |  | || |||
Db       3756 AAAAAGAAAAGAAGAGAAATTTTGGCATGCTGTAACATAATGAACCTGGAGGACATTATG 3697

Qy        734 CTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTTAATTA 785
              ||  ||  | | ||     |   | || | |  | | |     || || | |
Db       3696 CTCAGTGAAATAAGCCAGTCACAATATTACAAATACAATATGATTCTACTCA 3645


RESULT 12
AC144872/c
LOCUS       AC144872              173623 bp    DNA     linear   HTG 07-JUN-2003
DEFINITION  Pan troglodytes clone CH251-495I14, WORKING DRAFT SEQUENCE, 2
            ordered pieces.
ACCESSION   AC144872
VERSION     AC144872.2  GI:31455635
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 173623)
```

```
  AUTHORS   Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
            Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
            Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
            Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
            Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173623)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 173623)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 7, 2003 this sequence version replaced gi:31044285.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: esa
                 Center clone name: 495I14

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 173453 bases at least Q40
                 Consensus quality: 173497 bases at least Q30
                 Consensus quality: 173518 bases at least Q20
                 Insert size: 157000; agarose-fp
                 Insert size: 173523; sum-of-contigs
                 Quality coverage: 13.17x in Q20 bases; agarose-fp
                 Quality coverage: 11.92x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   132061: contig of 132061 bp in length
            *  132062   132161: gap of unknown length
            *  132162   173623: contig of 41462 bp in length.
FEATURES            Location/Qualifiers
     source         1. .173623
                    /organism="Pan troglodytes"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9598"
                    /clone="CH251-495I14"
                    /clone_lib="CH251"
     misc_feature   1. .132061
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:left"
     misc_feature   1. .65891
                    /note="clone overlaps with GenBank Accession Number
                    AC096851 clone RP43-19G1 (center project name clr)"
     misc_feature   71148. .173623
                    /note="clone overlaps with GenBank Accession Number
                    AC144747 clone CH251-160D7 (center project name esb)"
     gap            132062. .132161
                    /estimated_length=unknown
     misc_feature   132162. .173623
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:right"
     misc_feature   156029. .173623
                    /note="clone overlaps with GenBank Accession Number
                    AC096878 clone RP43-171C12 (center project name cls)"
ORIGIN

  Query Match            11.2%;  Score 224.8;  DB 14;  Length 173623;
  Best Local Similarity  73.3%;  Pred. No. 1.1e-36;
  Matches  302;  Conservative    0;  Mismatches  107;  Indels    3;  Gaps    1;

Qy        377 TTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGG 436
              ||||| ||    ||||| |||||||| |||| ||||  |||||||||||||| ||||||||
Db      75209 TTAAAAAGGAGAGAAATTGGCCAGGCTCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 75150

Qy        437 GAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGG 496
              |||||| |||||||| ||||||||||||||||||||||  ||||||||||||||||||||
Db      75149 GAGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGG 75090

Qy        497 TGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTA 556
              ||||||||||||||||||||||||    ||||||||||||||||||||||||| || |||||||
Db      75089 TGAAACCCTGTCTCTACTAAAAATAAAAAAATTAGCCAGGTGTGGTGGCAGGCGCCTGTA 75030

Qy        557 GTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTG 616
              |||||  ||||||  ||||| ||||||||  ||||||||  ||||| ||| ||||
Db      75029 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCCGGGAGGCGGAAGTTG 74970

Qy        617 CAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAG 673
              ||||  | |||||||   ||||||||||||||||||||  ||||| || ||||||||
Db      74969 CAGTGAGCCGAGATCGCGTCACTGCACTCCAGCCTGGTGACAGAACACGACTCCATCTCA 74910

Qy        674 ACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGAAATTCAT 733
              | ||  |||          || | |    |||   |        ||   | |  | || |||
Db      74909 AAAAAGAAAAGAAGAGAAATTTTGGCATGCTGTAACATAATGAACCTGGAGGACATTATG 74850

Qy        734 CTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTTAATTA 785
              ||  ||  | | |     |   | || | | | | |     || || | |
Db      74849 CTCAGTGAAATAAGCCAGTCACAATATTACAAATACAATATGATTCTACTCA 74798


RESULT 13
AC121247
LOCUS       AC121247               170787 bp    DNA     linear   PRI 19-NOV-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-3B7, complete sequence.
ACCESSION   AC121247 AC012124
VERSION     AC121247.2  GI:25101130
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 170787)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
```

```
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 170787)
  AUTHORS    Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-MAY-2002) Genome Center, University of Washington,
             Box 352145, Seattle, WA 98195, USA
REFERENCE    3  (bases 1 to 170787)
  AUTHORS    Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
             Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
             Haugen,E.D.
  TITLE      Direct Submission
  JOURNAL    Submitted (19-NOV-2002) Genome Center, University of Washington,
             Box 352145, Seattle, WA 98195, USA
COMMENT      On Nov 19, 2002 this sequence version replaced gi:20806311.
             -------------- Genome Center
                 Center: University of Washington Genome Center
                 Center Code: UWGC
                 Web site: http://www.genome.washington.edu
                 Contact: uwgchtgs@u.washington.edu
                 Drafting Center: BCM
             -------------- Project Information
                 Center project name: chr-3
                 Center clone name: RP11-3B7 (bc0105)
             ------------- Summary Statistics
                 Sequencing vector: plasmid; 47% of reads
                 Sequencing vector: unknown; 53% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 170463 bases at least Q40
                 Consensus quality: 170727 bases at least Q30
                 Consensus quality: 170782 bases at least Q20
                 Insert size: 170787; sum-of-contigs
                 Quality coverage: 9.4x in Q20 bases; sum-of-contigs
             --------------
             Overlapping Sequences:
                 5': Mapping in progress
                 3': RP11-804H15 AC104311
             --------------------
             Sequence Quality Assessment:
              This entry has been annotated with sequence quality
              estimates computed by the Phrap assembly program.
              All manually edited bases have been reduced to quality zero.
              Quality levels above 40 are expected to have less than
              1 error in 10,000 bp.
              Base-by-base quality values are not generally visible from the
              GenBank flat file format but are available as part
              of this entry's ASN.1 file.

             This sequence was finished as follows unless otherwise noted:
             all regions were either double-stranded or sequenced with an
             alternate chemistry or covered by high quality data (i.e., Phred
             quality >= 30); an attempt was made to resolve all sequencing
             problems, such as compressions and repeats; all regions were
             covered by at least one plasmid subclone or more than one M13
             subclone; and the assembly was confirmed by restriction digest.
             --------------------
             Sequence Validation:
              This sequence has been validated by Multiple Complete Digest
              fingerprinting. Comparison of the experimentally derived digest
              fragments with sequence-predicted fragments is given below.
              The electronically-digested sequence consists of both insert and
              vector, in order to accurately represent the entire circular BAC.
              Small fragments below a variable cutoff (approximately 400-800 bp)
              are not resolved in the fingerprint and hence do not appear
              in the table. There are no significant remaining discrepancies
              between the experimental and predicted values. Uniquely ordered
              fragments are separated by dashed lines.
                   HindIII                 EcoRI                 BglII
             ------------------  ------------------  ------------------
             SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt
             ---------  --------  ---------  --------  ---------  --------

             ---------            ---------            ---------
```

13802	13540	8696	8757	3877	3864
6382	6542	6	<800	2067	2046
512	<800	9988	9979	14311	14181
449	<800	8347	8314	2993	3053
5288	5288	1772	1747	3829	3864
660	<800	12907	12726	2535	2519
835	854	929	930	3072	3053
7851	7883	8700	8757	10768	10611
3365	3374	693	<800	1072	1062
39	<800	2756	2797	39	<800
4109	4061	14794	14646	5511	5565
189	<800	12996	12726	4883	4732
1697	1649	1165	1149	15360	15166
15845	15867	1309	1284	110	<800
941	948	3194	3202	3439	3511
5249	5288	7412	7507	1558	1529
8491	8664	15188	14646	769	778
3057	3069	13287	12726	3706	3864
4457	4471	3102	3087	4759	4732
1342	1306	4166	4243	212	<800
9716	9838	1452	1420	4701	4732
24363	24680	794	806	5330	5368
3172	3183	3383	3398	586	<800
10795	10782	2234	2257	1740	1716
4741	4684	3035	3087	37	<800
21198	21101	3377	3398	1272	1241
4653	4684	10041	9979	3479	3511
16285	15867	578	<800	678	<800
		10129	9979	5235	5176
		13053	12726	7875	8032
				22885	23142
				12447	12320
				5515	5565
				4945	5004
				7736	7788
				10152	10081

FEATURES Location/Qualifiers

```
     source          1. .170787
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-3B7"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN

  Query Match             11.2%;  Score 224.4;  DB 8;  Length 170787;
  Best Local Similarity   79.1%;  Pred. No. 1.3e-36;
  Matches  280;  Conservative    0;  Mismatches   71;  Indels    3;  Gaps    1;

Qy        334 ACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAAT 393
              ||||   |  ||    || | | |||  ||||    |||| ||   || |   | | |
Db     138920 ACAAGAGTGAAACTCCATCTCAGAAAAAAATAATAATAAAATAAATAATAATAATAATAA 138979

Qy        394 GGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGG 453
              |||||||| |||| ||||  |||| |||| ||| || ||||||||||  || |||| ||
Db     138980 TGGCCAGGCACAGTGGCTCACGCCTATAATTCCAGCATTTTGGGAGGCTGAGATGGGTGG 139039

Qy        454 ATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 513
              |||| ||||||||||||||||  ||||||||||||||||||||| ||||||||||||||
Db     139040 ATCAACTGAGGTCAGGAGTTTGTGACCAGCCTGGCCAACATGGTAAAACCCTGTCTCTAC 139099

Qy        514 TAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGA 573
              ||||||  |||||||||||||||  |||||||| || ||||||||||| ||||||||||
Db     139100 TAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGCGCCTGTAGTCCCAGCTACTCAGGA 139159

Qy        574 GGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---T 630
              || ||||| |||||||| |||||||||| |||||  |||||||||||| | || |||    |
Db     139160 GGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATTGT 139219

Qy        631 GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGC 684
              | |||||||||||||||||| |||||||||||||| | || | || |||| |
Db     139220 GCCACTGCACTCCAGCCTGGCGACAGAGCAAGACTCCGTCAAAAAAAAAAAATC 139273


RESULT 14
AC135506/c
LOCUS       AC135506               174363 bp    DNA     linear   PRI 23-APR-2003
DEFINITION  Homo sapiens chromosome 3 clone RP11-694I15, complete sequence.
ACCESSION   AC135506
VERSION     AC135506.3  GI:30061452
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 174363)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174363)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 174363)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2003) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   4  (bases 1 to 174363)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2003) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Apr 23, 2003 this sequence version replaced gi:28626666.
            ------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            ------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-694I15 (bc0806)
            ------------ Summary Statistics
                Sequencing vector: plasmid; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 174180 bases at least Q40
                Consensus quality: 174339 bases at least Q30
                Consensus quality: 174362 bases at least Q20
                Insert size: 174363; sum-of-contigs
                Quality coverage: 11.7x in Q20 bases; sum-of-contigs
            --------------.
            Overlapping Sequences:
                5':  RP13-131K19 AC137630, 2005-bp overlap
                3':  RP11-3B7 (UWGC:bc0105) AC121247, 33949-bp overlap
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
                  HindIII                 EcoRI                   BglII
            ------------------  ------------------  ------------------
            SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt
            ---------  --------  ---------  --------  ---------  --------

            ---------            ---------            ---------
                 3328      3315       8696      8841       7165      7121
            ---------            ---------            ---------
                 6382      6511          6      <800       2067      2034
            ---------            ---------            ---------
                  512      <800        929       930       5269      5408
            ---------            ---------            ---------
                  449      <800      12907     12649      10768     10490
            ---------            ---------            ---------
                11958     11682       1772      1743       3072      3010
            ---------            ---------            ---------
                 1697      1644       8347      8362       2535      2503
            ---------            ---------            ---------
                  189      <800       9988      9926       3829      3822
```

4109	4017	1091	1071	2993	3010
39	<800	14963	15082	20894	21024
3365	3315	1661	1625	9237	8982
7851	7831	11265	11328	8585	8982
835	881	10108	9926	793	<800
660	<800	15190	15082	5462	5408
8679	8754	189	<800	648	<800
14244	14033	3890	3920	316	<800
5769	5860	5708	5766	9849	9784
4655	4677	3039	3054	1211	1211
6520	6511	15804	16006	1878	1881
8654	8754	11492	11328	433	<800
8623	8754	15945	16006	3907	3999
8927	8754	51	<800	20	<800
3967	4017	2131	2148	4872	4903
2675	2684	4567	4606	6115	6167
16757	16905	779	778	4043	3999
7172	7286	11477	11328	2184	2182
5547	5584	302	<800	4448	4483
801	828	9721	9926	868	955
1353	1336	1044	1071	920	955
447	<800			2598	2650
7394	7510			8839	8982
3446	3448			645	<800
6132	6192			4731	4903
890	881			8855	8982
8111	8115			104	<800
50	<800			978	955
1028	1028			12456	12174
1797	1774			2400	2373
4203	4017			4969	4903
3847	3805			5012	4903
				175	<800
				5839	5788
				1080	1070

FEATURES Location/Qualifiers
source 1. .174363
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-694I15"
/clone_lib="RPCI human BAC library 11"
misc_feature 140036. .140049
/note="Single subclone region"
misc_feature 174350. .174363
/note="Single subclone region"

ORIGIN

Query Match 11.2%; Score 224.4; DB 8; Length 174363;
Best Local Similarity 79.1%; Pred. No. 1.3e-36;
Matches 280; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

```
Qy        334 ACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAAT 393
              ||||   |  ||    || | | |||  ||||    |||| ||   || |    | | |
Db     172282 ACAAGAGTGAAACTCCATCTCAGAAAAAAATAATAATAAAATAAATAATAATAATAATAA 172223

Qy        394 GGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGG 453
              |||||||| |||| ||||  |||| |||| ||| || ||||||||||  || |||| ||
Db     172222 TGGCCAGGCACAGTGGCTCACGCCTATAATTCCAGCATTTTGGGAGGCTGAGATGGGTGG 172163

Qy        454 ATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 513
              |||| ||||||||||||||||  ||||||||||||||||||||| ||||||||||||||
Db     172162 ATCAACTGAGGTCAGGAGTTTGTGACCAGCCTGGCCAACATGGTAAAACCCTGTCTCTAC 172103

Qy        514 TAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGA 573
              ||||||  |||||||||||||||  |||||||| || |||||||||||| ||||||||||
Db     172102 TAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGCGCCTGTAGTCCCAGCTACTCAGGA 172043

Qy        574 GGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---T 630
              || ||||| |||||||| |||||||||| |||||  |||||||||||| | || |||   |
Db     172042 GGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATTGT 171983

Qy        631 GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGC 684
              | |||||||||||||||||| |||||||||||||| | || | || |||| |
Db     171982 GCCACTGCACTCCAGCCTGGCGACAGAGCAAGACTCCGTCAAAAAAAAAAAATC 171929
```

RESULT 15
AC104311/c
LOCUS AC104311 185146 bp DNA linear HTG 21-JAN-2002
DEFINITION Homo sapiens chromosome UNK clone RP11-804H15, *** SEQUENCING IN PROGRESS ***, 45 unordered pieces.
ACCESSION AC104311
VERSION AC104311.2 GI:18252763
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 185146)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185146)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Jan 21, 2002 this sequence version replaced gi:17402818.

---------- Genome Center ----------
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
---------- Project Information ----------
Center project name: H_NH0804H15
---------- Summary Statistics ----------

```
Sequencing vector: M13; 73%
Sequencing vector: plasmid; 27%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 169877 bases at least Q40
Consensus quality: 173619 bases at least Q30
Consensus quality: 175935 bases at least Q20
-----------------------------------------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1178: contig of 1178 bp in length
*     1179     1278: gap of unknown length
*     1279     2877: contig of 1599 bp in length
*     2878     2977: gap of unknown length
*     2978     4291: contig of 1314 bp in length
*     4292     4391: gap of unknown length
*     4392     5675: contig of 1284 bp in length
*     5676     5775: gap of unknown length
*     5776     7129: contig of 1354 bp in length
*     7130     7229: gap of unknown length
*     7230     8368: contig of 1139 bp in length
*     8369     8468: gap of unknown length
*     8469    10108: contig of 1640 bp in length
*    10109    10208: gap of unknown length
*    10209    12010: contig of 1802 bp in length
*    12011    12110: gap of unknown length
*    12111    13979: contig of 1869 bp in length
*    13980    14079: gap of unknown length
*    14080    15503: contig of 1424 bp in length
*    15504    15603: gap of unknown length
*    15604    17771: contig of 2168 bp in length
*    17772    17871: gap of unknown length
*    17872    19412: contig of 1541 bp in length
*    19413    19512: gap of unknown length
*    19513    22030: contig of 2518 bp in length
*    22031    22130: gap of unknown length
*    22131    23419: contig of 1289 bp in length
*    23420    23519: gap of unknown length
*    23520    25351: contig of 1832 bp in length
*    25352    25451: gap of unknown length
*    25452    27518: contig of 2067 bp in length
*    27519    27618: gap of unknown length
*    27619    29406: contig of 1788 bp in length
*    29407    29506: gap of unknown length
*    29507    31125: contig of 1619 bp in length
*    31126    31225: gap of unknown length
*    31226    32948: contig of 1723 bp in length
*    32949    33048: gap of unknown length
*    33049    35570: contig of 2522 bp in length
*    35571    35670: gap of unknown length
*    35671    38933: contig of 3263 bp in length
*    38934    39033: gap of unknown length
*    39034    42434: contig of 3401 bp in length
*    42435    42534: gap of unknown length
*    42535    45155: contig of 2621 bp in length
*    45156    45255: gap of unknown length
*    45256    48463: contig of 3208 bp in length
*    48464    48563: gap of unknown length
*    48564    50927: contig of 2364 bp in length
*    50928    51027: gap of unknown length
*    51028    54065: contig of 3038 bp in length
*    54066    54165: gap of unknown length
*    54166    57491: contig of 3326 bp in length
*    57492    57591: gap of unknown length
*    57592    61643: contig of 4052 bp in length
*    61644    61743: gap of unknown length
*    61744    64595: contig of 2852 bp in length
*    64596    64695: gap of unknown length
*    64696    68934: contig of 4239 bp in length
*    68935    69034: gap of unknown length
*    69035    74299: contig of 5265 bp in length
*    74300    74399: gap of unknown length
*    74400    77602: contig of 3203 bp in length
*    77603    77702: gap of unknown length
*    77703    81677: contig of 3975 bp in length
*    81678    81777: gap of unknown length
*    81778    85979: contig of 4202 bp in length
*    85980    86079: gap of unknown length
*    86080    92090: contig of 6011 bp in length
*    92091    92190: gap of unknown length
*    92191    96442: contig of 4252 bp in length
*    96443    96542: gap of unknown length
*    96543   101168: contig of 4626 bp in length
*   101169   101268: gap of unknown length
*   101269   106935: contig of 5667 bp in length
*   106936   107035: gap of unknown length
*   107036   114293: contig of 7258 bp in length
*   114294   114393: gap of unknown length
*   114394   122291: contig of 7898 bp in length
*   122292   122391: gap of unknown length
*   122392   131013: contig of 8622 bp in length
*   131014   131113: gap of unknown length
*   131114   136091: contig of 4978 bp in length
*   136092   136191: gap of unknown length
*   136192   156515: contig of 20324 bp in length
*   156516   156615: gap of unknown length
*   156616   183997: contig of 27382 bp in length
*   183998   184097: gap of unknown length
*   184098   185146: contig of 1049 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185146
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="RP11-804H15"
     misc_feature    1. .1178
                     /note="assembly_name:Contig103"
     gap             1179. .1278
                     /estimated_length=unknown
     misc_feature    1279. .2877
                     /note="assembly_name:Contig104"
     gap             2878. .2977
                     /estimated_length=unknown
     misc_feature    2978. .4291
                     /note="assembly_name:Contig106"
     gap             4292. .4391
                     /estimated_length=unknown
     misc_feature    4392. .5675
                     /note="assembly_name:Contig108"
     gap             5676. .5775
                     /estimated_length=unknown
     misc_feature    5776. .7129
                     /note="assembly_name:Contig110"
     gap             7130. .7229
                     /estimated_length=unknown
     misc_feature    7230. .8368
                     /note="assembly_name:Contig111"
     gap             8369. .8468
                     /estimated_length=unknown
     misc_feature    8469. .10108
                     /note="assembly_name:Contig112"
     gap             10109. .10208
                     /estimated_length=unknown
     misc_feature    10209. .12010
                     /note="assembly_name:Contig114"
     gap             12011. .12110
                     /estimated_length=unknown
     misc_feature    12111. .13979
```

```
                     /note="assembly_name:Contig115"
     gap             13980. .14079
                     /estimated_length=unknown
     misc_feature    14080. .15503
                     /note="assembly_name:Contig116"
     gap             15504. .15603
                     /estimated_length=unknown
     misc_feature    15604. .17771
                     /note="assembly_name:Contig117"
     gap             17772. .17871
                     /estimated_length=unknown
     misc_feature    17872. .19412
                     /note="assembly_name:Contig118"
     gap             19413. .19512
                     /estimated_length=unknown
     misc_feature    19513. .22030
                     /note="assembly_name:Contig119"
     gap             22031. .22130
                     /estimated_length=unknown
     misc_feature    22131. .23419
                     /note="assembly_name:Contig120"
     gap             23420. .23519
                     /estimated_length=unknown
     misc_feature    23520. .25351

  Query Match             11.2%;  Score 224.4;  DB 14;  Length 185146;
  Best Local Similarity   79.1%;  Pred. No. 1.3e-36;
  Matches  280;  Conservative    0;  Mismatches   71;  Indels    3;  Gaps    1;

Qy        334 ACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAAT 393
              ||||   |  ||    || | | |||  ||||    |||| ||   || |    | | |
Db     120915 ACAAGAGTGAAACTCCATCTCAGAAAAAAATAATAATAAAATAAATAATAATAATAATAA 120856

Qy        394 GGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGG 453
              |||||||| |||| ||||  |||| |||| ||| || |||||||||   || |||| ||
Db     120855 TGGCCAGGCACAGTGGCTCACGCCTATAATTCCAGCATTTTGGGAGGCTGAGATGGGTGG 120796

Qy        454 ATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 513
              |||| ||||||||||||||||  |||||||||||||||||||| ||||||||||||||||
Db     120795 ATCAACTGAGGTCAGGAGTTTGTGACCAGCCTGGCCAACATGGTAAAACCCTGTCTCTAC 120736

Qy        514 TAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGA 573
              ||||||  ||||||||||||||||  |||||||| || ||||||||||||| ||||||||||
Db     120735 TAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGCGCCTGTAGTCCCAGCTACTCAGGA 120676

Qy        574 GGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---T 630
              || ||||| |||||||| |||||||||| |||||  |||||||||| | || |||    |
Db     120675 GGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATTGT 120616

Qy        631 GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGC 684
              | |||||||||||||||||| |||||||||||||| | || | || |||| |
Db     120615 GCCACTGCACTCCAGCCTGGCGACAGAGCAAGACTCCGTCAAAAAAAAAAAATC 120562


RESULT 16
AC073546/c
LOCUS       AC073546              141521 bp    DNA     linear   PRI 21-JUL-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-406P13, complete sequence.
ACCESSION   AC073546
VERSION     AC073546.5  GI:14993704
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 141521)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141521)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 141521)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 141521)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 21, 2001 this sequence version replaced gi:14550307.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1. .141521
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-406P13"
ORIGIN

  Query Match             11.2%;  Score 224.2;  DB 8;  Length 141521;
  Best Local Similarity   60.4%;  Pred. No. 1.5e-36;
  Matches  446;  Conservative    0;  Mismatches  278;  Indels   15;  Gaps    4;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||  ||  ||||  |||||  |||||||||||||||||||||||||||||| | ||
Db      91558 TGATCTGCCTGCCTCAGCCTCTGAAAGTGCTGGGATTACAGGCGTGAGCCACCGCACCCA 91499

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  | |||||| |||     | | |             | | |    | |||| |||
Db      91498 GCCATAAGTCTTTTTTAAAACACAAATCTG---------TTAGCCACACATCTAATTGTT 91448

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | | | ||  |   | | | ||    |    |  |||     || |     ||
Db      91447 AAAATATAATCCATAATCACCACCATAAGACAGGGCAAGGAAACTATATATAATCTGTAA 91388

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |   | |     || |||   | | |       |   |  | ||   ||| ||  || ||
Db      91387 TTATTTTTGTTGGTTTTTAACCTTTTTGCTTCAGATTTTTTAATCATTTTTAAATAGAAA 91328

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
                 ||  | |||  |||       |  |       | |      | |  || | |
Db      91327 CTTGGAAATTTACAGTACTGATTGCTATTCTGACATGCAGAATTTTCTGTGCACTAATGT 91268

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACT--ATTTAATAAAATTTTAAAA 358
              | | |  ||   |||  |  ||  |  | | |   |   |  |  ||   |
Db      91267 GGAAAGATTTCTTAGTACTAAAGGACATCCAGAAATATTTCAACAATCATTAAGTTTTCA 91208

Qy        359 ACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCT 418
                 ||  | || ||   ||        || ||||||||| |  || ||||  ||||
Db      91207 TGACTTTCATATTAAATACAGAAATATAAGAAATTGGGCCAGGTGAGGTGGCTCAGGCCT 91148

Qy        419 GTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGA 478
              |||||||| |||||||||||||| ||||| | |||||| |||||||||||||| ||||
Db      91147 GTAATCCCAGCACTTTGGGAGGCCGAGGTGAGAGGATCACTTGAGGTCAGGAGTTCAAGA 91088

Qy        479 CCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTG 538
              |||||||||||||||| ||||||| |  |||||||||||  ||||||||||  | ||
Db      91087 CCAGCCTGGCCAACATAGTGAAACTCCATCTCTACTAAAAATACAAAAATTAGCTGGATG 91028

Qy        539 TGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAA 598
              |||||||||||||||||||||||| |||||||||||  ||||| ||||||||||||||||
Db      91027 TGGTGGCATGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAA 90968

Qy        599 CCTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAAC 654
```

```
            || ||||||  |||||||||||  | || ||   ||| |||||||||  ||||||  ||||||
Db      90967 CCCAGGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCACTGCACTTCAGCCTGGGCAAC 90908

Qy        655 AGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTAT 714
              ||||||||||  |   |     | ||  |||     | || ||     || || ||   ||  ||
Db      90907 AGAGCAAGACCCTGTCTGAAAAATTAAAAAAATTAATTTAAAAAATTGGGTGTTATATTT 90848

Qy        715 CTTCATTTAAGAAATTCAT 733
                | |  |    || | ||
Db      90847 AGTAACATTTTTAAGTAAT 90829


RESULT 17
AC026465/c
LOCUS       AC026465               142838 bp    DNA     linear   HTG 26-SEP-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-354J24, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC026465
VERSION     AC026465.2  GI:10305170
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 142838)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142838)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 26, 2000 this sequence version replaced gi:7280277.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 542541
            Center clone name: RPCI-11_354J24
            -------
            Summary Statistics
            Consensus quality: 123712 bases at least Q40
            Consensus quality: 135403 bases at least Q30
            Consensus quality: 137509 bases at least Q20
            Estimated insert size: 194240; agarose-fp estimation
            Estimated insert size: 140738; sum-of-contigs estimation
            Quality coverage:  3.28 in Q20 bases; agarose-fp estimation
            Quality coverage:  4.53 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1225: contig of 1225 bp in length
            *     1226     1325: gap of unknown length
            *     1326     2352: contig of 1027 bp in length
            *     2353     2452: gap of unknown length
            *     2453     3799: contig of 1347 bp in length
            *     3800     3899: gap of unknown length
            *     3900     5202: contig of 1303 bp in length
            *     5203     5302: gap of unknown length
            *     5303     6648: contig of 1346 bp in length
            *     6649     6748: gap of unknown length
            *     6749     9853: contig of 3105 bp in length
            *     9854     9953: gap of unknown length
            *     9954    11640: contig of 1687 bp in length
            *    11641    11740: gap of unknown length
            *    11741    13774: contig of 2034 bp in length
            *    13775    13874: gap of unknown length
            *    13875    15777: contig of 1903 bp in length
            *    15778    15877: gap of unknown length
            *    15878    17736: contig of 1859 bp in length
            *    17737    17836: gap of unknown length
            *    17837    20854: contig of 3018 bp in length
            *    20855    20954: gap of unknown length
            *    20955    24018: contig of 3064 bp in length
            *    24019    24118: gap of unknown length
            *    24119    29675: contig of 5557 bp in length
            *    29676    29775: gap of unknown length
            *    29776    36388: contig of 6613 bp in length
            *    36389    36488: gap of unknown length
            *    36489    43469: contig of 6981 bp in length
            *    43470    43569: gap of unknown length
            *    43570    51109: contig of 7540 bp in length
            *    51110    51209: gap of unknown length
            *    51210    61408: contig of 10199 bp in length
            *    61409    61508: gap of unknown length
            *    61509    71666: contig of 10158 bp in length
            *    71667    71766: gap of unknown length
            *    71767    85149: contig of 13383 bp in length
            *    85150    85249: gap of unknown length
            *    85250   100523: contig of 15274 bp in length
            *   100524   100623: gap of unknown length
            *   100624   119207: contig of 18584 bp in length
            *   119208   119307: gap of unknown length
            *   119308   142838: contig of 23531 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142838
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-354J24"
                     /clone_lib="RPCI human BAC library 11"
     gap             1226. .1325
                     /estimated_length=unknown
     gap             2353. .2452
                     /estimated_length=unknown
     gap             3800. .3899
                     /estimated_length=unknown
     gap             5203. .5302
                     /estimated_length=unknown
     gap             6649. .6748
                     /estimated_length=unknown
     gap             9854. .9953
                     /estimated_length=unknown
     gap             11641. .11740
                     /estimated_length=unknown
     gap             13775. .13874
                     /estimated_length=unknown
     gap             15778. .15877
                     /estimated_length=unknown
     gap             17737. .17836
                     /estimated_length=unknown
     gap             20855. .20954
                     /estimated_length=unknown
     gap             24019. .24118
                     /estimated_length=unknown
     gap             29676. .29775
                     /estimated_length=unknown
     gap             36389. .36488
                     /estimated_length=unknown
     gap             43470. .43569
                     /estimated_length=unknown
     gap             51110. .51209
                     /estimated_length=unknown
     gap             61409. .61508
                     /estimated_length=unknown
     gap             71667. .71766
```

```
                     /estimated_length=unknown
     gap             85150. .85249
                     /estimated_length=unknown
     gap             100524. .100623
                     /estimated_length=unknown
     gap             119208. .119307
                     /estimated_length=unknown
ORIGIN

  Query Match           11.2%;  Score 224.2;  DB 14;  Length 142838;
  Best Local Similarity   60.4%;  Pred. No. 1.5e-36;
  Matches  446;  Conservative    0;  Mismatches  278;  Indels   15;  Gaps    4;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||  || ||||  |||||  ||||||||||||||||||||||||||||| | ||
Db     108011 TGATCTGCCTGCCTCAGCCTCTGAAAGTGCTGGGATTACAGGCGTGAGCCACCGCACCCA 107952

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  | |||||| |||    |  |  |          | |  |   | |||| |||
Db     107951 GCCATAAGTCTTTTTTAAAACACAAATCTG---------TTAGCCACACATCTAATTGTT 107901

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | | | ||  |      | | | ||     |      |  |||       || |     ||
Db     107900 AAAATATAATCCATAATCACCACCATAAGACAGGGCAAGGAAACTATATATAATCTGTAA 107841

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |   | |      || |||   | | |      |   |  | ||   ||| ||  || ||
Db     107840 TTATTTTTGTTGGTTTTTAACCTTTTTGCTTCAGATTTTTTAATCATTTTTAAATAGAAA 107781

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
                ||  | |||  |||       |  |       | |      | |  || | |
Db     107780 CTTGGAAATTTACAGTACTGATTGCTATTCTGACATGCAGAATTTTCTGTGCACTAATGT 107721

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACT--ATTTAATAAAATTTTAAAA 358
                | |  |   ||   |||  |  ||  |  | |  |   |  |  |  ||  |
Db     107720 GGAAAGATTTCTTAGTACTAAAGGACATCCAGAAATATTTCAACAATCATTAAGTTTTCA 107661

Qy        359 ACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCT 418
                | ||  | || ||   ||       || ||||||||| |  || ||||  ||||
Db     107660 TGACTTTCATATTAAATACAGAAATATAAGAAATTGGGCCAGGTGAGGTGGCTCAGGCCT 107601

Qy        419 GTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGA 478
              ||||||||| |||||||||||||| ||||| | ||||||| ||||||||||||| ||||
Db     107600 GTAATCCCAGCACTTTGGGAGGCCGAGGTGAGAGGATCACTTGAGGTCAGGAGTTCAAGA 107541

Qy        479 CCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTG 538
              |||||||||||||||  |||||||  |  ||||||||||||  ||||||||||  | ||
Db     107540 CCAGCCTGGCCAACATAGTGAAACTCCATCTCTACTAAAAATACAAAAATTAGCTGGATG 107481

Qy        539 TGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAA 598
              |||||||||||||||||||||||| |||||||||||| ||||| ||||||||||||||||
Db     107480 TGGTGGCATGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAA 107421

Qy        599 CCTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAAC 654
              || |||||  |||||||||||| | || ||   ||| ||||||||| |||||| ||||||
Db     107420 CCCAGGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCACTGCACTTCAGCCTGGGCAAC 107361

Qy        655 AGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTAT 714
              ||||||||||  |   |     | ||  |||     | || ||    || || ||  ||  | |
Db     107360 AGAGCAAGACCCTGTCTGAAAAATTAAAAAAATTAATTTAAAAAATTGGGTGTTATATTT 107301

Qy        715 CTTCATTTAAGAAATTCAT 733
                | |  |    || | ||
Db     107300 AGTAACATTTTTAAGTAAT 107282


RESULT 18
AC022445
LOCUS       AC022445               180807 bp    DNA     linear   PRI 18-DEC-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-489L13, complete sequence.
ACCESSION   AC022445
VERSION     AC022445.6  GI:17921207
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 180807)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180807)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 180807)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 180807)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 18, 2001 this sequence version replaced gi:15799584.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.7% of Sequence;
            Estimated Total Number of Errors is 0.4.
FEATURES             Location/Qualifiers
     source          1. .180807
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-489L13"
ORIGIN

  Query Match           11.2%;  Score 224.2;  DB 8;  Length 180807;
  Best Local Similarity   60.4%;  Pred. No. 1.5e-36;
  Matches  446;  Conservative    0;  Mismatches  278;  Indels   15;  Gaps    4;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||  || ||||  |||||  ||||||||||||||||||||||||||||| | ||
Db      23342 TGATCTGCCTGCCTCAGCCTCTGAAAGTGCTGGGATTACAGGCGTGAGCCACCGCACCCA 23401

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  | |||||| |||    |  |  |          | |  |   | |||| |||
Db      23402 GCCATAAGTCTTTTTTAAAACACAAATCTG---------TTAGCCACACATCTAATTGTT 23452

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | | | ||  |      | | | ||     |      |  |||       || |     ||
Db      23453 AAAATATAATCCATAATCACCACCATAAGACAGGGCAAGGAAACTATATATAATCTGTAA 23512

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |   | |      || |||   | | |      |   |  | ||   ||| ||  || ||
Db      23513 TTATTTTTGTTGGTTTTTAACCTTTTTGCTTCAGATTTTTTAATCATTTTTAAATAGAAA 23572

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
                ||  | |||  |||       |  |       | |      | |  || | |
Db      23573 CTTGGAAATTTACAGTACTGATTGCTATTCTGACATGCAGAATTTTCTGTGCACTAATGT 23632

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACT--ATTTAATAAAATTTTAAAA 358
                | |  |   ||   |||  |  ||  |  | |  |   |  |  |  ||  |
Db      23633 GGAAAGATTTCTTAGTACTAAAGGACATCCAGAAATATTTCAACAATCATTAAGTTTTCA 23692

Qy        359 ACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCT 418
                | ||  | || ||   ||       || ||||||||| |  || ||||  ||||
Db      23693 TGACTTTCATATTAAATACAGAAATATAAGAAATTGGGCCAGGTGAGGTGGCTCAGGCCT 23752

Qy        419 GTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGA 478
```

```
              ||||||||| |||||||||||||| ||||| | ||||||| ||||||||||||||| ||||
Db      23753 GTAATCCCAGCACTTTGGGAGGCCGAGGTGAGAGGATCACTTGAGGTCAGGAGTTCAAGA 23812

Qy        479 CCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTG 538
              |||||||||||||||| ||||||| |  ||||||||||||  |||||||||| | ||
Db      23813 CCAGCCTGGCCAACATAGTGAAACTCCATCTCTACTAAAAATACAAAAATTAGCTGGATG 23872

Qy        539 TGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAA 598
              |||||||||||||||||||||||| |||||||||||  ||||| ||||||||||||||||
Db      23873 TGGTGGCATGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAA 23932

Qy        599 CCTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAAC 654
              || |||||||  |||||||||||| | || ||   ||| |||||||||| |||||| ||||||
Db      23933 CCCAGGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCACTGCACTTCAGCCTGGGCAAC 23992

Qy        655 AGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTAT 714
              |||||||||| |  |     | ||  |||     | || ||     || || ||  ||  | |
Db      23993 AGAGCAAGACCCTGTCTGAAAAATTAAAAAAATTAATTTAAAAAATTGGGTGTTATATTT 24052

Qy        715 CTTCATTTAAGAAATTCAT 733
                | |  |    || | ||
Db      24053 AGTAACATTTTTAAGTAAT 24071


RESULT 19
AC159033/c
LOCUS       AC159033               165467 bp    DNA     linear   HTG 24-MAR-2005
DEFINITION  Pan troglodytes chromosome UNKNOWN clone CH251-339G2, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC159033
VERSION     AC159033.1  GI:61741063
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 165467)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Pan troglodytes clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165467)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2005) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: C_AB0339G02
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 160219 bases at least Q40
            Consensus quality: 161271 bases at least Q30
            Consensus quality: 161983 bases at least Q20
            ------------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1371: contig of 1371 bp in length
            *     1372     1471: gap of unknown length
            *     1472     3143: contig of 1672 bp in length
            *     3144     3243: gap of unknown length
            *     3244     4584: contig of 1341 bp in length
            *     4585     4684: gap of unknown length
            *     4685     6388: contig of 1704 bp in length
            *     6389     6488: gap of unknown length
            *     6489    33801: contig of 27313 bp in length
            *    33802    33901: gap of unknown length
            *    33902    59513: contig of 25612 bp in length
            *    59514    59613: gap of unknown length
            *    59614    84765: contig of 25152 bp in length
            *    84766    84865: gap of unknown length
            *    84866   165467: contig of 80602 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165467
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="UNKNOWN"
                     /clone="CH251-339G2"
     misc_feature    1. .1371
                     /note="assembly_name:Contig10"
     gap             1372. .1471
                     /estimated_length=unknown
     misc_feature    1472. .3143
                     /note="assembly_name:Contig35"
     gap             3144. .3243
                     /estimated_length=unknown
     misc_feature    3244. .4584
                     /note="assembly_name:Contig36"
     gap             4585. .4684
                     /estimated_length=unknown
     misc_feature    4685. .6388
                     /note="assembly_name:Contig42"
     gap             6389. .6488
                     /estimated_length=unknown
     misc_feature    6489. .33801
                     /note="assembly_name:Contig43
                     clone_end:SP6
                     vector_side:left"
     gap             33802. .33901
                     /estimated_length=unknown
     misc_feature    33902. .59513
                     /note="assembly_name:Contig44"
     gap             59514. .59613
                     /estimated_length=unknown
     misc_feature    59614. .84765
                     /note="assembly_name:Contig45"
     gap             84766. .84865
                     /estimated_length=unknown
     misc_feature    84866. .165467
                     /note="assembly_name:Contig46
                     clone_end:T7
                     vector_side:left"
ORIGIN

  Query Match             11.2%;  Score 223.2;  DB 14;  Length 165467;
  Best Local Similarity   85.1%;  Pred. No. 2.4e-36;
  Matches  262;  Conservative    0;  Mismatches   43;  Indels    3;  Gaps    1;

Qy        380 AAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAG 439
              || | |||||   | |||||||| |||| |||| ||| |||||||||| ||||||||||
Db      38467 AAAACCTTTGTCTTCGGCCAGGCACAGTGGCTCATGCGTGTAATCCCAGCACTTTGGGAG 38408

Qy        440 GCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGA 499
              | |||||||| |||||||||| |||||||||| ||||||||||||||||||||||||||
Db      38407 GTCAAGGTGGGTGGATCACCTGAAGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGA 38348

Qy        500 AACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTC 559
              |||||||||||||||||||  ||||||||||||||  |||||||  || |||||| ||
Db      38347 AACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGCTTGGTGGCGGGCCCCTGTAATC 38288
```

```
Qy        560 CCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAG 619
              ||| |||||||||||| ||||| |||||||| ||||||||| |||||| ||||||||||
Db      38287 CCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGTTGCAG 38228

Qy        620 TAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACA 676
              | | |||||||   | ||||||||||||||||||  |||||| |||||||   ||| | |
Db      38227 TGAGCCGAGATCACGCCACTGCACTCCAGCCTGGAGACAGAGTAAGACTCTGAAAAAAAA 38168

Qy        677 ACAAAAGC 684
              |  ||| |
Db      38167 AACAAAAC 38160
```

RESULT 20
AC079177
LOCUS AC079177 155533 bp DNA linear PRI 15-MAR-2002
DEFINITION Homo sapiens Xp BAC RP11-147B14 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC079177
VERSION AC079177.21 GI:13489128
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155533)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155533)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 155533)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 155533)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-APR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 155533)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 6 (bases 1 to 155533)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Mar 30, 2001 this sequence version replaced gi:13375595.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

---------- Summary Statistics ----------

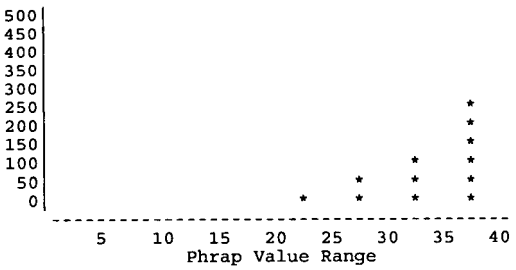
Contig length:	155533
Phrap values in estimate:	154305
Average error rate (BCM-Phrap estimate):	2.34429e-06
Fraction of Phrap values less than 40 :	0.00316905
Number of consensus changing edits:	10

Number of N's in consensus : 0

---------- Consensus changing edits ----------

Position	Original+Context	Edited+Context
36096	gataccacgc(n)tgacccccat	gataccacgc(c)tgacccccat
36145	aaccccattt(g)cccatatttc	aaccccattt(c)cccatatttc
36460	atgccccgag(n)caagaaacta	atgccccgag(t)caagaaacta
45749	gcctcttggg(n)nnnagcaatt	gcctcttggg(t)tccagcaatt
45750	cctcttgggn(n)nnagcaattc	cctcttgggt(t)ccagcaattc
45751	ctcttgggnn(n)nagcaattct	ctcttgggtt(c)cagcaattct
45752	tcttgggnnn(n)agcaattctc	tcttgggttc(c)agcaattctc
66973	atacatttca(n)gacagtttct	atacatttca(t)gacagtttct
153681	ttgactgccc(n)taagtctcta	ttgactgccc(c)taagtctcta
153695	gtctctagga(n)acattgggga	gtctctagga(t)acattgggga

--------- Distribution of Quality < 40 Bases ----------



Version: 1.01 qxfo.

FEATURES	Location/Qualifiers
source	1. .155533
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="X"
	/clone="RP11-147B14"
misc_feature	complement(1. .2008)
	/note="Overlaps bases 2008. .1 of clone AC090632"
	/function="Overlaps with adjacent clone AC090632"
repeat_region	665. .891
	/rpt_family="MIR"
repeat_region	1145. .1175
	/rpt_family="AT_rich"
repeat_region	1232. .1533
	/rpt_family="AluSx"
repeat_region	complement(1606. .2542)
	/rpt_family="L1MA4A"
repeat_region	2694. .2764
	/rpt_family="U6"
repeat_region	3606. .3634
	/rpt_family="AT_rich"
repeat_region	3662. .3933
	/rpt_family="AluSx"
repeat_region	4622. .5049
	/rpt_family="L2"
repeat_region	5095. .5135
	/rpt_family="(TG)n"
repeat_region	5136. .5172
	/rpt_family="(GA)n"
repeat_region	5229. .5619
	/rpt_family="MLT2CB"
repeat_region	5637. .5700

Query Match 11.1%; Score 222.2; DB 8; Length 155533;
Best Local Similarity 61.1%; Pred. No. 3.9e-36;

Matches 453; Conservative 0; Mismatches 268; Indels 20; Gaps 5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| |||||||| ||||||||||||||| |
Db      42742 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 42801

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |      ||| |   | |   | | |
Db      42802 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 42861

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |      ||| || |  || | | |    |   |       | |
Db      42862 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 42921

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| |||  |     |       |  |  |||| ||
Db      42922 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 42981

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||              ||  |  | ||  |  |  ||  |   ||  |   | || ||
Db      42982 TTAG-------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 43028

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | |  | |||  | |  |  |  |    |
Db      43029 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 43086

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              |    |       | | ||  |  || ||   ||||||| || || ||| ||||||
Db      43087 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 43146

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| ||||||||||||||||||||||||||||||||||||||| ||||
Db      43147 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 43206

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||  ||| || ||||||||||||||  |||||||||||||||  |
Db      43207 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 43266

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||||
Db      43267 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 43326

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||| | || ||||   | ||||||||||||||| ||  |||
Db      43327 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 43386

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |      ||      ||| | |||   | ||
Db      43387 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 43446

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |   |||  ||| | |
Db      43447 GGAAGAGCTGAATCTCAGCAT 43467
```

RESULT 21
AC023271/c

LOCUS	AC023271	106499 bp	DNA	linear	PRI 16-APR-2005

DEFINITION Homo sapiens BAC clone RP11-96E8 from 2, complete sequence.
ACCESSION AC023271
VERSION AC023271.11 GI:14589672
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 106499)
AUTHORS Cordum,H., Kang,K. and Dignan,G.
TITLE The sequence of Homo sapiens BAC clone RP11-96E8
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 106499)
AUTHORS Waterston,R.H.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 106499)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 106499)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 106499)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:14165364.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0096E08
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-544H14, 2000 bp overlap;
            the clone sequenced to the right is RP11-33N4, 2000 bp overlap.
            Actual end of this clone is at base position 24907 of RP11-33N4.

            Data from AC080084 was used to finish RP11-96E8.
FEATURES              Location/Qualifiers
     source           1. .106499
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="2"
                      /clone="RP11-96E8"
                      /clone_lib="RPCI-11"
     misc_feature     1118. .1759
                      /note="CpG_island (%GC=68.5, o/e=0.74, #CpGs=58)"
```

```
ORIGIN

  Query Match             11.1%;  Score 222;  DB 8;  Length 106499;
  Best Local Similarity   61.7%;  Pred. No. 4.6e-36;
  Matches  489;  Conservative    0;  Mismatches  290;  Indels   13;  Gaps    8;

Qy          2 GATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTGG 61
              ||| | || |||| |||||| |   ||||||||||||||||     |||||||||| |
Db      63875 GATTCTCCTGCCTCAGCCTCCGGAGTAGCTGGGATTACAGGCACCTGCCACCACGCCCAG 63816

Qy         62 TCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAA----TCTGTCTTCTATTT 117
               | ||| | | |||| || | ||||||  || | |  |      || | || | |
Db      63815 -CTAATTTTGTATTTTTAGTAGAGACCTCAGGTGACCCACCCGCCTCAGCCTCCCAAAG 63757

Qy        118 GACAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACT 177
               | ||  ||   ||  | |    |||   | | | | |  | | | | |||     ||
Db      63756 TGCTGAGATTACAGGCATGAGCCACCACACCCGGCCCCTTGT-CAGGTCTTTATGCTGCT 63698

Qy        178 TAGTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAG 237
              | ||| | |||  ||  | | |      | |||      || |   |   ||| |
Db      63697 TTTTCTTTGTCATATGCATGTAT-AGCCTCTAAAAAATAAAATAAAATAATTTATAATTT 63639

Qy        238 GAATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAG 297
               || | |   | |  | || |    | | |   |||| |    ||| |  | | ||
Db      63638 TAAAAGAAAGGTTACTATTGGTACCAGAATTTTAGAATATTAGTCCTGGAGTGGACCCAG 63579

Qy        298 TAGTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAA 357
               |  | |  | |  ||   ||      || |   | |  | |   |  | || |||||
Db      63578 GAAATGTTTTTGTCCAACTCTTTCATTTTCTAGTTGATTAGT-GATGGTCAATTTTTACT 63520

Qy        358 AACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCC 417
              ||  |      | | | | |||    |      |  | |||  |||||||| |||| ||||
Db      63519 AATGAGAGGTGTAATAGTCTTAGTAAAGAGACATCTAGGCTGGGCGCAGTGGCTCATGCC 63460

Qy        418 TGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAG 477
              |||||||||| |||||||||||||| |||  |||||||||||||||||||||||||  ||
Db      63459 TGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCGGATCACCTGAGGTCAGGAGTTCGAG 63400

Qy        478 ACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGT 537
              |||||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||
Db      63399 ACCAGCCTGGCCAACATGGTGAAATCCTGTCTCTACTAAAAATGCAAAAATTAGCCAGGT 63340

Qy        538 GTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGA 597
              ||||||||  |||||||| ||||| ||||| ||||| ||||| ||||||||  |||||
Db      63339 GTGGTGGCGGGCACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCTCTTGA 63280

Qy        598 ACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAA 653
              |||  ||||||  |||||||||| | | ||||   || |||||||  |||||||  |||||
Db      63279 ACCCAGGAGGCAGAGGTTGCAGTGAGCTGAGATCGTGCCACTGCATTCCAGCCTGGGCAA 63220

Qy        654 CAGAGCAAGACTCCAT-AAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCT 712
              |||||||| ||| |||    | | || ||||      ||| |   ||     |  | |
Db      63219 CAGAGCAACACTTCATCTCAAAAAAAAAAAAAAAAAAAAAGTACATAGTACATATATGTAC 63160

Qy        713 ATCTTCATTTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAG 772
               || |||   |    ||   |||| |   || ||   | || ||||| |    | ||
Db      63159 TCCTGCATCATACTTTTTTTGCTTTTTAACTTAATACATTCTGGACATTACTACATCTTT 63100

Qy        773 CAATTTTTAATT 784
              | | ||||  ||
Db      63099 CTACTTTTTTTT 63088


RESULT 22
AC025750
LOCUS       AC025750              151367 bp    DNA     linear   PRI 30-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-804P20 from 2, complete sequence.
ACCESSION   AC025750
VERSION     AC025750.10  GI:18098549
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Hominidae; Homo.
REFERENCE 1  (bases 1 to 151367)
  AUTHORS Cotton,M., Doebber,A. and Tomlinson,C.
  TITLE   The sequence of Homo sapiens BAC clone RP11-804P20
  JOURNAL Unpublished (2001)
REFERENCE 2  (bases 1 to 151367)
  AUTHORS Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL Submitted (13-MAR-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 3  (bases 1 to 151367)
  AUTHORS Waterston,R.
  TITLE   Direct Submission
  JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4  (bases 1 to 151367)
  AUTHORS Wilson,R.K.
  TITLE   Direct Submission
  JOURNAL Submitted (30-APR-2005) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
COMMENT   On Jan 9, 2002 this sequence version replaced gi:13431263.
          -------------- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu
              Contact: submissions@watson.wustl.edu
          -------------- Summary Statistics
              Center project name: H_NH0804P20
          --------------.

          NOTICE:

          This sequence was finished as follows unless otherwise noted:
          all regions were double stranded, sequenced with an alternate
          chemistry, or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by sequence
          from more than one subclone; and the assembly was confirmed by
          restriction digest.

          MAPPING INFORMATION:
          Mapping information for this clone was provided by Dr. Wes Warren,
          Department of Genetics, Washington University, St. Louis MO. For
          additional information about the map position of this sequence, see
          http://genome.wustl.edu

          SOURCE INFORMATION:
          The RPCI-11 human BAC library was made from the blood of one male
          donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
          Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
          approach for construction of bacterial artificial chromosome
          libraries. Genomics 51:1-8. The clone may be obtained either from
          Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
          and coworkers at http://www.chori.org
          VECTOR:  pBACe3.6

          NEIGHBORING SEQUENCE INFORMATION:
          The clone sequenced to the left is RP11-299C5, 2000 bp overlap; the
          clone sequenced to the right is RP11-729M8.  Actual start of this
          clone is at base position 157736 of RP11-299C5; actual end is at
          base position 151367 of RP11-804P20.

          Data from  AC074375 and AC006038 was used to finish this clone,
          AC025750. Polymorphisms have been identified between AC025750 and
          AC074375. The sequence from 40567 to 41115, from 40991 to 41115 are
          covered only by pcr products from clone DNA. There is an unresolved
          homopolymeric runs between 39229 and 39243. There is an unresolved
          region between 132476 and 132479.
FEATURES            Location/Qualifiers
     source         1. .151367
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="2"
                    /clone="RP11-804P20"
                    /clone_lib="RPCI-11"
     gene           complement(5917. .57997)
                    /gene="KCNG3"
     mRNA           complement(join(5917. .8479,56737. .57997))
                    /gene="KCNG3"
     CDS            complement(join(7834. .8479,56737. .57401))
                    /gene="KCNG3"
                    /note="Homo sapiens potassium voltage-gated channel,
                    subfamily G, member 3 (KCNG3), transcript variant 1,
                    mRNA.; H_NH0804P20.1
                    This gene was based on gi(27436991)"
                    /codon_start=1
                    /product="unknown"
                    /protein_id="AAY24039.1"
                    /db_xref="GI:62988651"
                    /translation="MTFGRSGAASVVLNVGGARYSLSRELLKDFPLRRVSRLHGCRSE
                    RDVLEVCDDYDRERNEYFFDRHSEAFGFILLYVRGHGKLRFAPRMCELSFYNEMIYWG
                    LEGAHLEYCCQRRLDDRMSDTYTFYSADEPGVLGRDEARPGGAEAAPSRRWLERMRRT
                    FEEPTSSLAAQILASVSVVFVIVSMVVLCASTLPDWRNAAADNRSLDDRSRYSAGPGR
                    EPSGIIEAICIGWFTAECIVRFIVSKNKCEFVKRPLNIIDLLAITPYYISVLMTVFTG
                    ENSQLQRAGVTLRVLRMMRIFWVIKLARHFIGLQTLGLTLKRCYREMVMLLVFICVAM
                    AIFSALSQLLEHGLDLETSNKDFTSIPAACWWVIISMTTVGYGDMYPITVPGRILGGV
                    CVVSGIVLLALPITFIYHSFVQCYHELKFRSARYSRSLSTEFLN"
     misc_feature   56545. .58926
                    /note="CpG_island (%GC=71.6, o/e=0.94, #CpGs=297)"
     misc_feature   132370. .132990
                    /note="CpG_island (%GC=77.0, o/e=0.87, #CpGs=67)"
     gene           132585. .143100
                    /gene="tmp_locus_6"
     mRNA           join(132585. .132673,134337. .134404,143007. .143100)
                    /gene="tmp_locus_6"
     CDS            join(132646. .132673,134337. .134404,143007. .>143100)
                    /gene="tmp_locus_6"
                    /note="Homo sapiens, Similar to metastasis associated 1,
                    clone MGC:4778 IMAGE:3542706, mRNA, complete cds.;
                    H_NH0804P20.2
                    This gene was based on gi(13278950)
                    Continues as H_NH0729M08.1"
                    /codon_start=1
                    /product="unknown"
                    /protein_id="AAY24040.1"
                    /db_xref="GI:62988652"
                    /translation="MAANMYRVGDYVYFENSSSNPYLIRRIEELNKTASGNVEAKVVC
                    FYRRRDISNTLIMLADKHA"
ORIGIN

  Query Match           11.1%;  Score 222;  DB 8;  Length 151367;
  Best Local Similarity  60.6%;  Pred. No. 4.3e-36;
  Matches  417;  Conservative    0;  Mismatches  265;  Indels    6;  Gaps    3;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||| || |||| ||||||||||||||||||||||| |||| ||||||| |||| |
Db      30890 TGATCCGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTATAGGCAGGAGCCACTGCGCCCG 30949

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | | ||      | || | | || | | || |     | ||      | | | | | |
Db      30950 GCCAAAAACAGATCTTTTATAAAAACATAAAAATATATTAAACCACTCCATGAAGATAAA 31009

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              |       |  || ||||    |  | |||   |     |          ||   ||
Db      31010 TGTGAAAATTTAGGTGAAATGGACAATTTCCCAGAAAAATGAATCAGACCAAAATTTTCA 31069

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | |  |  |      | |      |  ||| ||||     |  |     |
Db      31070 GTTTTAGAATAACTGGAAAAACTGAACGGTCCTACAATCATGAAAAAAATTGAATCTACA 31129

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
                || |      | |||      |    | || |   |    |||||
```

```
Db      31130 CTTTGACATCTATTCACCAAAAGATGCATACCGGGTCCAGATGATTTTACAGGCAAGTTC 31189

Qy        301 TAGCTGCTATTGAAGAT--TTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAA 358
              || |   ||   |  ||   |||    |||||  || |||  | |   |||   ||||||
Db      31190 TATCATATACCAAGCATGAATAATTCCTATTTTATATAACACTGTCCCAAATAATTAAAA 31249

Qy        359 ACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCT 418
              |   |  |||  |     || || |  |  |  |  |||||| ||||| | |||  ||||
Db      31250 AAGCAGTCACCAACTCATTTTAAAAAGTCAGTGAAAGGCCAGTCGCAGCAACTCACGCCT 31309

Qy        419 GTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGA 478
              ||||||||| |||||||||||||||||| ||| |||||| |||||||||||||  ||||
Db      31310 GTAATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGATCATCTGAGGTCAGGAGTGCAAGA 31369

Qy        479 CCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTG 538
              |||||||||||||||||||||| ||  ||||||||||||  ||||||| |||||||
Db      31370 CCAGCCTGGCCAACATGGTGAAAACCCATCTCTACTAAAAATACAAAAATGAGCCAGGCA 31429

Qy        539 TGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAA 598
              ||||||||    |||||| ||||| |||||||||||| ||||| ||||||||  ||||||
Db      31430 TGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAA 31489

Qy        599 CCTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAAC 654
              || |||||||||||||||||| | | |||   ||| ||||||||||||||| ||  ||
Db      31490 CCCAGGAGGTGGAGGTTGCAGTGAACTGAGATCATGCCACTGCACTCCAGCCTGGGTGAC 31549

Qy        655 AGAGCAAGACTCCATAAAGACAACAAAA 682
              ||||  ||||||  |    | || ||||
Db      31550 AGAGTGAGACTCTGTTTCAAAAAAAAA 31577
```

```
RESULT 23
AP003164
LOCUS       AP003164               90289 bp    DNA     linear   HTG 16-AUG-2001
DEFINITION  Homo sapiens chromosome 1 clone dJ756A17 map 1p36.2-p36, ***
            SEQUENCING IN PROGRESS ***, 6 unordered pieces.
ACCESSION   AP003164
VERSION     AP003164.1  GI:15208266
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Machida,T., Ohira,M., Morohashi,A., Mihara,M., Furuta,S., Soeda,E.
            and Nakagawara,A.
  TITLE     Homo sapiens 90,289 genomic DNA of 1p36.2-p36
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 90289)
  AUTHORS   Machida,T., Ohira,M., Morohashi,A., Mihara,M., Furuta,S., Soeda,E.
            and Nakagawara,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) Akira Nakagawara, Chiba Cancer Center
            Research Institute, Division of Biochemistry; 666-2 Nitona,
            Chuoh-ku, Chiba, Chiba 260-8717, Japan
            (E-mail:akiranak@chiba-cc.pref.chiba.jp, Tel:81-43-264-5431,
            Fax:81-43-265-4459)
COMMENT     Additional author information
            Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A.,
            Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
            The Institute of Physical and Chemical Research (RIKEN) Genomic
            Sciences Center (GSC)
            1-7-22 Suehiro-chou,Tsurumi-ku,Yokohama,Kanagawa, JAPAN zip:
            230-0045
            phone: 81-45-503-9111, fax: 81-45-503-9170
            e-mail: hattori@gsc.riken.go.jp
            --------------
            NOTE: This is a 'working draft' sequence. It currently consists of
            6 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                1     13927 contig of  13927 bp in length
            14028     22029 contig of   8002 bp in length
            22130     56617 contig of  34488 bp in length
            56718     66346 contig of   9629 bp in length
            66447     76543 contig of  10097 bp in length
            76643     90289 contig of  13646 bp in length
            NOTE: This is a 'working draft' sequence. It currently consists of
            6 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                1     13927 contig of  13927 bp in length
            14028     22029 contig of   8002 bp in length
            22130     56617 contig of  34488 bp in length
            56718     66346 contig of   9629 bp in length
            66447     76543 contig of  10097 bp in length
            76643     90289 contig of  13646 bp in length.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    13927: contig of 13927 bp in length
            *   13928    14027: gap of 100 bp
            *   14028    22029: contig of 8002 bp in length
            *   22030    22129: gap of 100 bp
            *   22130    56617: contig of 34488 bp in length
            *   56618    56717: gap of 100 bp
            *   56718    66346: contig of 9629 bp in length
            *   66347    66446: gap of 100 bp
            *   66447    76543: contig of 10097 bp in length
            *   76544    76643: gap of 100 bp
            *   76644    90289: contig of 13646 bp in length.
FEATURES            Location/Qualifiers
     source         1. .90289
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="1"
                    /map="1p36.2-p36"
                    /clone="dJ756A17"
ORIGIN

  Query Match           11.1%;  Score 221.2;  DB 14;  Length 90289;
  Best Local Similarity   83.8%;  Pred. No. 7e-36;
  Matches  263;  Conservative    0;  Mismatches   48;  Indels    3;  Gaps    1;

Qy        388 TGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGT 447
              |||| | |||||||||||||| || ||  |||||||||||| ||||||||| |||| ||||
Db      20511 TGAAGTCGGCCAGGCGCAGTGGCCCCCACCTGTAATCCCAGCACTTTGGGTGGCCGAGGT 20570

Qy        448 GGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGT 507
              |||||||||| |||||||||||||| |||||||||||||||||||| ||||||||  |
Db      20571 GGGCGGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGATGAAACCCCAT 20630

Qy        508 CTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTAC 567
              |||||||||||  ||||||| || | || |||||| |||||||||||| ||| | ||||
Db      20631 CTCTACTAAAAATACAAAAATGAGTCGGGCGTGGTGCCATGCACCTGTAATCCTAGCTAC 20690

Qy        568 TCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGA 627
              |||||||| ||||| |||||||||||||||||| ||||| |||||||||||| | ||||
Db      20691 TCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGA 20750

Qy        628 GAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGC 684
              |||   | |||||||||||||||||  ||||||| |||||||||    | || ||||
Db      20751 GATGGCGCCACTGCACTCCAGCCTGGAGACAGAGCGAGACTCCATCTCAAAAAAAAAAAA 20810
```

```
Qy        685 TTTGAAATTGTGTA 698
                |||||  || |
Db      20811 AAATAAATTCTGAA 20824


RESULT 24
AX039602
LOCUS       AX039602              121162 bp    DNA     linear   PAT 18-NOV-2000
DEFINITION  Sequence 1 from Patent WO0063375.
ACCESSION   AX039602
VERSION     AX039602.1  GI:11229631
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Bougueleret,L., Dufaure-Gare,I. and Grel,P.
  TITLE     Dna encoding a kinesin-like protein (hklp) comprising biallelic ma
            rkers
  JOURNAL   Patent: WO 0063375-A 1 26-OCT-2000;
            GENSET (FR)
FEATURES             Location/Qualifiers
     source          1. .121162
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     exon            4844. .4920
                     /note="exon"
     primer_bind     7041. .7060
                     /note="12-809.pu"
     exon            7058. .7237
                     /note="exon"
     primer_bind     7140. .7158
                     /note="12-809-119.mis"
     misc_binding    7147. .7171
                     /bound_moiety="12-809-119.probe"
     variation       7159
                     /note="12-809-119 : polymorphic base G or C"
     primer_bind     7160. .7178
                     /note="12-809-119.mis complement"
     primer_bind     7545. .7565
                     /note="12-809.rp complement"
     exon            10469. .10534
                     /note="exon"
     exon            15940. .16118
                     /note="exon"
     primer_bind     16004. .16024
                     /note="12-805.rp"
     primer_bind     16350. .16368
                     /note="12-805-115.mis"
     misc_binding    16357. .16381
                     /bound_moiety="12-805-115.probe"
     variation       16369
                     /note="12-805-115 : polymorphic base A or G"
     primer_bind     16370. .16388
                     /note="12-805-115.mis complement"
     primer_bind     16464. .16483
                     /note="12-805.pu complement"
     exon            16712. .16823
                     /note="exon"
     exon            20054. .20131
                     /note="exon"
     exon            20793. .20858
                     /note="exon"
     exon            21558. .21575
                     /note="exon"
     exon            23971. .24046
                     /note="exon"
     exon            24864. .24942
                     /note="exon"
     primer_bind     26232. .26252
                     /note="12-790.rp"
     primer_bind     26291. .26309
                     /note="12-790-396.mis"
     misc_binding    26298. .26322
                     /bound_moiety="12-790-396.probe"
     variation       26310
                     /note="12-790-396 : polymorphic base A or G"
     primer_bind     26311. .26329
                     /note="12-790-396.mis complement"
     exon            26529. .26671
                     /note="exon"
     primer_bind     26685. .26705
                     /note="12-790.pu complement"
     exon            27640. .27681
                     /note="exon"
     exon            30863. .31074
                     /note="exon"
     primer_bind     30902. .30922
                     /note="12-791.pu"
     primer_bind     31093. .31111
                     /note="12-791-211.mis"
     misc_binding    31100. .31124
                     /bound_moiety="12-791-211.probe"
     variation       31112
                     /note="12-791-211 : polymorphic base A or G"
     primer_bind     31113. .31131
                     /note="12-791-211.mis complement"
     primer_bind     31282. .31301
                     /note="12-791.rp complement"
     primer_bind     33476. .33496
                     /note="12-803.rp"
     primer_bind     33789. .33807
                     /note="12-803-125.mis"
     misc_binding    33796. .33820
                     /bound_moiety="12-803-125.probe"
     variation       33808
                     /note="12-803-125 : polymorphic base A or T"
     primer_bind     33809. .33827
                     /note="12-803-125.mis complement"
     primer_bind     33912. .33932
                     /note="12-803.pu complement"
     primer_bind     33934. .33953
                     /note="99-33040.pu"
     primer_bind     34236. .34254
                     /note="99-33040-321.mis"
     misc_binding    34243. .34267
                     /bound_moiety="99-33040-321.probe"
     variation       34255
                     /note="99-33040-321 : polymorphic base C or T"
     primer_bind     34256. .34274
                     /note="99-33040-321.mis complement"
     primer_bind     34364. .34383
                     /note="99-33040.rp complement"
     primer_bind     34918. .34938
                     /note="12-810.rp"
     primer_bind     35274. .35292
                     /note="12-810-77.mis"
     misc_binding    35281. .35305
                     /bound_moiety="12-810-77.probe"
     variation       35293
                     /note="12-810-77 : polymorphic base A or G"
     primer_bind     35294. .35312
                     /note="12-810-77.mis complement"
     primer_bind     35350. .35369
                     /note="12-810.pu complement"
     exon            39625. .39704
                     /note="exon"
     exon            40590. .40665
                     /note="exon"
     exon            43630. .43709
                     /note="exon"
     exon            44204. .44310
```

```
                /note="exon"
     exon       45126. .45209
                /note="exon"
     exon       45441. .45621
                /note="exon"
     primer_bind 45465. .45485
                /note="12-787.rp"
     exon       45718. .45790
                /note="exon"
     primer_bind 45873. .45891
                /note="12-787-103.mis"
     misc_binding 45880. .45904
                /bound_moiety="12-787-103.probe"
     variation  45892
                /note="12-787-103 : polymorphic base A or G"
     primer_bind 45893. .45911
                /note="12-787-103.mis complement"
     primer_bind 45974. .45993
                /note="12-787.pu complement"
     primer_bind 54361. .54381
                /note="12-793.rp"
     primer_bind 54478. .54496
                /note="12-793-383.mis"
     misc_binding 54485. .54509
                /bound_moiety="12-793-383.probe"
     variation  54497
                /note="12-793-383 : polymorphic base G or T"
     primer_bind 54498. .54516
                /note="12-793-383.mis complement"
     primer_bind 54861. .54879
                /note="12-793.pu complement"
     primer_bind 56365. .56385
                /note="12-792.rp"
     primer_bind 56563. .56581
                /note="12-792-233.mis"
     misc_binding 56570. .56594
                /bound_moiety="12-792-233.probe"
     variation  56582
                /note="12-792-233 : polymorphic base A or G"
     primer_bind 56583. .56601
                /note="12-792-233.mis complement"
     primer_bind 56793. .56813
                /note="12-792.pu complement"
     primer_bind 60111. .60130
                /note="99-41009.rp"
     primer_bind 60317. .60335
                /note="99-41009-244.mis"
     misc_binding 60324. .60348
                /bound_moiety="99-41009-244.probe"
     variation  60336
                /note="99-41009-244 : polymorphic base A or G"
     primer_bind 60337. .60355
                /note="99-41009-244.mis complement"
     primer_bind 60450. .60468
                /note="99-41009-111.mis"
     misc_binding 60457. .60481
                /bound_moiety="99-41009-111.probe"
     variation  60469
                /note="99-41009-111 : polymorphic base C or T"
     primer_bind 60470. .60488
                /note="99-41009-111.mis complement"
     primer_bind 60562. .60580

  Query Match             11.1%;  Score 221.2;  DB 6;  Length 121162;
  Best Local Similarity   83.8%;  Pred. No. 6.6e-36;
  Matches  263;  Conservative    0;  Mismatches   48;  Indels    3;  Gaps    1;

Qy        388 TGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGT 447
              |||| | |||||||||||||| || ||  |||||||||||| ||||||||| |||| ||||
Db      25335 TGAAGTCGGCCAGGCGCAGTGGCCCCCACCTGTAATCCCAGCACTTTGGGTGGCCGAGGT 25394

Qy        448 GGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGT 507
              |||||||||||| |||||||||||||| |||||||||||||||||||| ||||||||  |
Db      25395 GGGCGGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGATGAAACCCCAT 25454

Qy        508 CTCTACTAAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTAC 567
              |||||||||||  ||||||| || | || |||||| |||||||||||| ||| | ||||
Db      25455 CTCTACTAAAAATACAAAAATGAGTCGGGCGTGGTGCCATGCACCTGTAATCCTAGCTAC 25514

Qy        568 TCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGA 627
              |||||||| ||||| |||||||||||||||||||| ||||| |||||||||||| | ||||
Db      25515 TCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGA 25574

Qy        628 GAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGC 684
              |||   | |||||||||||||||||  ||||||| |||||||||     | || ||||
Db      25575 GATGGCGCCACTGCACTCCAGCCTGGAGACAGAGCGAGACTCCATCTCAAAAAAAAAAAA 25634

Qy        685 TTTGAAATTGTGTA 698
                |||||  || |
Db      25635 AAATAAATTCTGAA 25648


RESULT 25
AL358013
LOCUS       AL358013              171612 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-420G9 on chromosome 1 Contains
            the 5' end of the KIF1B gene for kinesin family member 1B and a CpG
            island, complete sequence.
ACCESSION   AL358013
VERSION     AL358013.19  GI:17017702
KEYWORDS    HTG; KIF1B.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 171612)
  AUTHORS   Dunn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 20, 2001 this sequence version replaced gi:16973843.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-420G9 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .171612
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-420G9"
```

```
                /clone_lib="RPCI-11.2"
misc_feature    2000
                /note="Clone_right_end: RP4-575L21"
gene            join(25447. .25520,46892. .47076,70889. .70965,73135. .73314,
                76547. .76612,82022. .82200,82794. .82905,86144. .86221,
                86883. .86948,90070. .90145,90963. .91041,92628. .92770,
                97042. .97175,105724. .105803,106689. .106764,109728. .109807,
                110302. .110408,111224. .111307,111539. .111719,
                111816. .111888,134685. .134778,136351. .136499,
                138526. .138704,139400. .139537,140753. .141001,
                149162. .149280,151299. .151384,151716. .151845,
                152013. .152175,154411. .154501,156692. .156810,
                157874. .157929,160487. .160595,162403. .162469,
                163291. .163375,167273. .167378,AL139424.21:5959. .6073,
                AL139424.21:6722. .6855,AL139424.21:8313. .8374,
                AL139424.21:10130. .10275,AL139424.21:10439. .10678,
                AL139424.21:13497. .13568,AL139424.21:16171. .16292,
                AL139424.21:19346. .19495,AL139424.21:19884. .20076,
                AL139424.21:20285. .20403,AL139424.21:21575. .22967)
                /gene="KIF1B"
                /locus_tag="RP4-736L20.1-002"
mRNA            join(25447. .25520,46892. .47076,70889. .70965,73135. .73314,
                76547. .76612,82022. .82200,82794. .82905,86144. .86221,
                86883. .86948,90070. .90145,90963. .91041,92628. .92770,
                97042. .97175,105724. .105803,106689. .106764,109728. .109807,
                110302. .110408,111224. .111307,111539. .111719,
                111816. .111888,134685. .134778,136351. .136499,
                138526. .138704,139400. .139537,140753. .141001,
                149162. .149280,151299. .151384,151716. .151845,
                152013. .152175,154411. .154501,156692. .156810,
                157874. .157929,160487. .160595,162403. .162469,
                163291. .163375,167273. .167378,AL139424.21:5959. .6073,
                AL139424.21:6722. .6855,AL139424.21:8313. .8374,
                AL139424.21:10130. .10275,AL139424.21:10439. .10678,
                AL139424.21:13497. .13568,AL139424.21:16171. .16292,
                AL139424.21:19346. .19495,AL139424.21:19884. .20076,
                AL139424.21:20285. .20403,AL139424.21:21575. .22967)
                /gene="KIF1B"
                /locus_tag="RP4-736L20.1-002"
                /product="kinesin family member 1B"
                /note="match: cDNAs: Em:AF257176.1"
gene            join(25447. .25520,46892. .47076,70889. .70965,73135. .73314,
                76547. .76612,82022. .82200,82794. .82905,86144. .86221,
                86883. .86948,90070. .90145,90963. .91041,92628. .92770,
                97042. .97175,105724. .105803,106689. .106764,109728. .109807,
                110302. .110408,111224. .111307,111539. .111719,
                111816. .111888,117805. .123239)
                /gene="KIF1B"
                /locus_tag="RP4-736L20.1-003"
mRNA            join(25447. .25520,46892. .47076,70889. .70965,73135. .73314,
                76547. .76612,82022. .82200,82794. .82905,86144. .86221,
                86883. .86948,90070. .90145,90963. .91041,92628. .92770,
                97042. .97175,105724. .105803,106689. .106764,109728. .109807,
                110302. .110408,111224. .111307,111539. .111719,
                111816. .111888,117805. .123239)
                /gene="KIF1B"
                /locus_tag="RP4-736L20.1-003"
                /product="kinesin family member 1B"
                /note="match: ESTs: Em:AL135264.1 Em:AU131505.1
                Em:BF338098.1 Em:BG258423.1 Em:BM460386.1 Em:CD517715.1
                Em:CF121324.1
                match: cDNAs: Em:AY139835.1"
gene            join(26258. .26380,46892. .47076,70889. .70965,73135. .73314,
                76547. .76612,82022. .82200,82794. .82905,86144. .86221,
                86883. .86948,87656. .87673,90070. .90145,90963. .91041,
                92628. .92770,93739. .93780,96964. .97175,105724. .105803,
                106689. .106764,109728. .109807,110302. .110408,
                111224. .111307,111539. .111719,111816. .111888,
                134685. .134778,136351. .136499,138526. .138704,
                139400. .139537,140753. .141001,149162. .149280,
                151299. .151384,151716. .151845,152013. .152175,
                154411. .154501,156692. .156810,157874. .157929,
                160487. .160595,162403. .162469,163291. .163375,
                167273. .167378,AL139424.21:5959. .6073,
                AL139424.21:6722. .6855,AL139424.21:8313. .8374,
                AL139424.21:10130. .10275,AL139424.21:10439. .10678,
                AL139424.21:13497. .13568,AL139424.21:16171. .16292,
                AL139424.21:19346. .19495,AL139424.21:19884. .20076,
                AL139424.21:20285. .20403,AL139424.21:21575. .26633)
                /gene="KIF1B"
                /locus_tag="RP4-736L20.1-001"
mRNA            join(26258. .26380,46892. .47076,70889. .70965,73135. .73314,
                76547. .76612,82022. .82200,82794. .82905,86144. .86221,
                86883. .86948,87656. .87673,90070. .90145,90963. .91041,
                92628. .92770,93739. .93780,96964. .97175,105724. .105803,
                106689. .106764,109728. .109807,110302. .110408,
                111224. .111307,111539. .111719,111816. .111888,
                134685. .134778,136351. .136499,138526. .138704,
                139400. .139537,140753. .141001,149162. .149280,
                151299. .151384,151716. .151845,152013. .152175,
                154411. .154501,156692. .156810,157874. .157929,
                160487. .160595,162403. .162469,163291. .163375,
                167273. .167378,AL139424.21:5959. .6073,
                AL139424.21:6722. .6855,AL139424.21:8313. .8374,
                AL139424.21:10130. .10275,AL139424.21:10439. .10678,
                AL139424.21:13497. .13568,AL139424.21:16171. .16292,
                AL139424.21:19346. .19495,AL139424.21:19884. .20076,
                AL139424.21:20285. .20403,AL139424.21:21575. .26633)
                /gene="KIF1B"
                /locus_tag="RP4-736L20.1-001"
                /product="kinesin family member 1B"
                /note="match: ESTs: Em:AI082256.1 Em:BG702590.1
                Em:BI601300.1 Em:BM716071.1 Em:BQ427458.1 Em:BQ772488.1
                match: cDNAs: Em:AB011163.2"
gene            join(45406. .45639,46892. .47076,70889. .70965,73135. .73314,
                76547. .76612,82022. .82200,82794. .82905,86144. .86221,
                86883. .86948,90070. .90145,90963. .91041,92628. .92770,
                97042. .97175,105724. .105803,106689. .106764,109728. .109807,
                110302. .110408,111224. .111307,111539. .111719,
                111816. .111888,117805. .121399)
                /gene="KIF1B"
                /locus_tag="RP4-736L20.1-004"
mRNA            join(45406. .45639,46892. .47076,70889. .70965,73135. .73314,
                76547. .76612,82022. .82200,82794. .82905,86144. .86221,
                86883. .86948,90070. .90145,90963. .91041,92628. .92770,
                97042. .97175,105724. .105803,106689. .106764,109728. .109807,
                110302. .110408,111224. .111307,111539. .111719,
                111816. .111888,117805. .121399)
                /gene="KIF1B"
                /locus_tag="RP4-736L20.1-004"
                /product="kinesin family member 1B"
                /note="match: cDNAs: Em:AB040881.2 Em:AY043362.1"
gene            join(46892. .47076,70889. .70965,73135. .73314,76547. .76612,
                82022. .82200,82794. .82905,86144. .86221,86883. .86948,
                87656. .87673,90070. .90145,90963. .91041,92628. .92770,
                93739. .93780,96964. .97175,105724. .105803,106689. .106764,
                109728. .109807,110302. .110408,111224. .111307,
                111539. .111719,111816. .111888,134685. .134778,
                136351. .136499,138526. .138704,139400. .139537,
                140753. .141001,149162. .149280,151299. .151384,
                151716. .151845,152013. .152175,154411. .154501,
                156692. .156810,157874. .157929,160487. .160595,
                162403. .162469,163291. .163375,167273. .167378,
                AL139424.21:5959. .6073,AL139424.21:6722. .6855,
                AL139424.21:8313. .8374,AL139424.21:10130. .10275,
                AL139424.21:10439. .10678,AL139424.21:13497. .13568,
                AL139424.21:16171. .16292,AL139424.21:19346. .19495,
                AL139424.21:19884. .20076,AL139424.21:20285. .20403,
                AL139424.21:23369. .26627)
                /gene="KIF1B"
                /locus_tag="RP4-736L20.1-005"
mRNA            join(46892. .47076,70889. .70965,73135. .73314,76547. .76612,
                82022. .82200,82794. .82905,86144. .86221,86883. .86948,
                87656. .87673,90070. .90145,90963. .91041,92628. .92770,
                93739. .93780,96964. .97175,105724. .105803,106689. .106764,
                109728. .109807,110302. .110408,111224. .111307,
```

```
                    111539. .111719,111816. .111888,134685. .134778,
                    136351. .136499,138526. .138704,139400. .139537,

  Query Match             11.1%;  Score 221.2;  DB 8;  Length 171612;
  Best Local Similarity   83.8%;  Pred. No. 6.2e-36;
  Matches  263;  Conservative    0;  Mismatches   48;  Indels    3;  Gaps    1;

Qy        388 TGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGT 447
              |||| | |||||||||||||| || ||  |||||||||||| |||||||| |||| ||||
Db      91434 TGAAGTCGGCCAGGCGCAGTGGCCCCCACCTGTAATCCCAGCACTTTGGGTGGCCGAGGT 91493

Qy        448 GGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGT 507
              |||||||||||| |||||||||||||| |||||||||||||||||| ||||||||  |
Db      91494 GGGCGGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGATGAAACCCCAT 91553

Qy        508 CTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTAC 567
              ||||||||||||  ||||||| || | || |||||| |||||||||||| ||| | ||||
Db      91554 CTCTACTAAAAATACAAAAATGAGTCGGGCGTGGTGCCATGCACCTGTAATCCTAGCTAC 91613

Qy        568 TCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGA 627
              |||||||| ||||| |||||||||||||||||| ||||| |||||||||||| | ||||
Db      91614 TCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGA 91673

Qy        628 GAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGC 684
              |||   | |||||||||||||||||  ||||||| |||||||||    | || ||||
Db      91674 GATGGCGCCACTGCACTCCAGCCTGGAGACAGAGCGAGACTCCATCTCAAAAAAAAAAAA 91733

Qy        685 TTTGAAATTGTGTA 698
                  ||||| || |
Db      91734 AAATAAATTCTGAA 91747


RESULT 26
AC137058/c
LOCUS       AC137058              185623 bp    DNA     linear   HTG 21-FEB-2003
DEFINITION  Papio anubis clone RP41-126M5, WORKING DRAFT SEQUENCE, 10 ordered
            pieces.
ACCESSION   AC137058
VERSION     AC137058.3  GI:28460766
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
  ORGANISM  Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Cercopithecidae; Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 185623)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185623)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 185623)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Feb 21, 2003 this sequence version replaced gi:27476124.
            ---------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ---------- Project Information
                 Center project name: dtj
                 Center clone name: 126M05

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ---------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 183092 bases at least Q40
                 Consensus quality: 183851 bases at least Q30
                 Consensus quality: 184434 bases at least Q20
                 Insert size: 160000; agarose-fp
                 Insert size: 184723; sum-of-contigs
                 Quality coverage: 14.13x in Q20 bases; agarose-fp
                 Quality coverage: 12.24x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     9984: contig of 9984 bp in length
            *     9985    10084: gap of unknown length
            *    10085    49805: contig of 39721 bp in length
            *    49806    49905: gap of unknown length
            *    49906    66936: contig of 17031 bp in length
            *    66937    67036: gap of unknown length
            *    67037    83604: contig of 16568 bp in length
            *    83605    83704: gap of unknown length
            *    83705   108246: contig of 24542 bp in length
            *   108247   108346: gap of unknown length
            *   108347   112021: contig of 3675 bp in length
            *   112022   112121: gap of unknown length
            *   112122   134817: contig of 22696 bp in length
            *   134818   134917: gap of unknown length
            *   134918   147290: contig of 12373 bp in length
            *   147291   147390: gap of unknown length
            *   147391   185337: contig of 37947 bp in length
            *   185338   185437: gap of unknown length
            *   185438   185623: contig of 186 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185623
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="RP41-126M5"
                     /clone_lib="RP41"
     misc_feature    1. .15263
                     /note="clone overlaps with GenBank Accession Number
                     AC130273 clone RP41-30706 (center project name dsu)"
     misc_feature    1. .9984
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     gap             9985. .10084
                     /estimated_length=unknown
```

```
     misc_feature    10085. .49805
                     /note="assembly_fragment"
     gap             49806. .49905
                     /estimated_length=unknown
     misc_feature    49906. .66936
                     /note="assembly_fragment"
     gap             66937. .67036
                     /estimated_length=unknown
     misc_feature    67037. .83604
                     /note="assembly_fragment"
     gap             83605. .83704
                     /estimated_length=unknown
     misc_feature    83705. .108246
                     /note="assembly_fragment"
     gap             108247. .108346
                     /estimated_length=unknown
     misc_feature    108347. .112021
                     /note="assembly_fragment"
     gap             112022. .112121
                     /estimated_length=unknown
     misc_feature    112122. .134817
                     /note="assembly_fragment"
     gap             134818. .134917
                     /estimated_length=unknown
     misc_feature    134918. .147290
                     /note="assembly_fragment"
     gap             147291. .147390
                     /estimated_length=unknown
     misc_feature    147391. .185337
                     /note="assembly_fragment"
     gap             185338. .185437
                     /estimated_length=unknown
     misc_feature    185438. .185623
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             11.0%;  Score 221;  DB 14;  Length 185623;
  Best Local Similarity   62.7%;  Pred. No. 6.7e-36;
  Matches  432;  Conservative    0;  Mismatches  240;  Indels   17;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||| || | || ||||||||||||||||||||||||||||| |||||||| |||| |
Db      36364 TGATCCGCCCGTCTCGGCCTCCCAAAGTGCTGGGATTACAGGCTTGAGCCACCGCGCCCG 36305

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  ||   || ||| ||| || | |  |||          ||     | | |   |||
Db      36304 GCCCTATTATTTGATTTTTTAAAAAAAATACA--------CAACTACACATTTTGCTGAG 36253

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | |||| |    |   ||| ||    | ||||  |||    |   ||   |  |||
Db      36252 GTATTTTGTTACCTTTCTATCTCATACTTACCACAGACTTTCTGTTTTTCTTTCTCTTTA 36193

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              ||||| |    |    |||| | |    | ||       |||   || |    | |
Db      36192 TCTCTTTGTATTTGAATTCTTCCTCCCAATGGGTTTTCAATTTGGGTTCCTCAAGGCCTC 36133

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              | ||  | |  |       |  ||| | | |    | | ||  | |||| |    | ||
Db      36132 TGAAGATTTATATGCCTGGAAATATATTTTATCCCCTTACATTTGAATTTTATTTGGCAG 36073

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |  | |  |  || ||| ||| | | ||  ||       || ||    |    |
Db      36072 GATATACATTCTAAAATTAAAGTGATTTTCCTTCTGTGCTTTAAACTTCACCCCACTGTT 36013

Qy        361 TAATACACTTAAATT---ATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCC 417
              |  | || |||          | ||||  | | |    |||||||  | || | ||  |||
Db      36012 TCTTGCATTTAGTACTGCTGTCAAGAAATCTTACGCCGGCCAGGTCCGGTGGTTCACGCC 35953

Qy        418 TGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAG 477
              |||||||||| |||||||||||||||||||| ||||| | |||||||||||||   ||
Db      35952 TGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGTGGATCTCTTGAGGTCAGGAGTTCCAG 35893

Qy        478 ACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGT 537
              |||||||||||||||||||||||||||  ||||||||||||   ||||| ||||| ||
Db      35892 ACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAA--TAAAAAATAGCCGGGC 35835

Qy        538 GTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGA 597
              |||||||   |||||| ||   ||| |||||||||||| ||||| ||||||||  |||||
Db      35834 GTGGTGGTGGGCACCTTTAAGTCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGA 35775

Qy        598 ACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAA 653
              ||| ||||||  |||| || ||| | || | ||   | ||||||||||||||| |||||
Db      35774 ACCCAGGAGGCAGAGGCTGTAGTGAGCCAAAATCACGCCACTGCACTCCAGCCTGGGCAA 35715

Qy        654 CAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||| ||||||| |       || ||||
Db      35714 CAGAGCTAGACTCCGTCTCAGAAAAAAAA 35686


RESULT 27
AC020603
LOCUS       AC020603               202317 bp    DNA     linear   PRI 09-MAY-2001
DEFINITION  Homo sapiens BAC clone RP11-534K13 from 8, complete sequence.
ACCESSION   AC020603
VERSION     AC020603.4  GI:9454643
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 202317)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 202317)
  AUTHORS   Kruchowski,S., Stoneking,T., Gregory,S. and Markovic,C.
  TITLE     The sequence of Homo sapiens BAC clone RP11-534K13
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 202317)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 202317)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 202317)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 202317)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 26, 2000 this sequence version replaced gi:7631048.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0534K13
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
```

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RP11-534K13; actual end is at base position 202317 of RP11-534K13.

FEATURES Location/Qualifiers

source	1. .202317
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="8"
	/map="8"
	/clone="RP11-534K13"
	/clone_lib="RPCI-11"
repeat_region	1392. .1595
	/rpt_family="MIR"
misc_feature	1794. .1880
	/note="similar to EST AW426876 (NID:g6954823)"
repeat_region	2198. .2412
	/rpt_family="MIR"
repeat_region	2418. .2498
	/rpt_family="L1"
repeat_region	2499. .2787
	/rpt_family="Alu"
repeat_region	2788. .2851
	/rpt_family="L1"
repeat_region	3054. .3144
	/rpt_family="L1"
repeat_region	3147. .3307
	/rpt_family="L1"
repeat_region	3887. .4243
	/rpt_family="L1"
repeat_region	4260. .4560
	/rpt_family="L1"
repeat_region	5055. .5774
	/rpt_family="L1"
repeat_region	5775. .6075
	/rpt_family="Alu"
repeat_region	6076. .6175
	/rpt_family="L1"
repeat_region	6282. .6520
	/rpt_family="L1"
repeat_region	6566. .7138
	/rpt_family="L1"
repeat_region	7354. .7831
	/rpt_family="L1"
repeat_region	9181. .9397
	/rpt_family="MER1_type"
repeat_region	9859. .9948
	/rpt_family="L2"
repeat_region	10576. .10822
	/rpt_family="L1"
repeat_region	10844. .10995
	/rpt_family="L1"
repeat_region	10996. .11301
	/rpt_family="Alu"
repeat_region	11718. .12803
	/rpt_family="L1"
repeat_region	12804. .13745
	/rpt_family="L1"
repeat_region	14052. .14330
	/rpt_family="MaLR"
repeat_region	14331. .14627
	/rpt_family="Alu"
repeat_region	14628. .14817
	/rpt_family="MaLR"
repeat_region	15075. .15300
	/rpt_family="MER1_type"
repeat_region	15419. .15575
	/rpt_family="Alu"
repeat_region	16346. .16654
	/rpt_family="Alu"
repeat_region	17018. .17230
	/rpt_family="CR1"
repeat_region	17549. .17730
	/rpt_family="MIR"
repeat_region	17875. .18176
	/rpt_family="Alu"
repeat_region	18187. .18631
	/rpt_family="ERVL"
repeat_region	18996. .19320
	/rpt_family="Alu"
repeat_region	19330. .19393
	/rpt_family="MIR"
repeat_region	19394. .19700
	/rpt_family="Alu"
repeat_region	19701. .19766
	/rpt_family="MIR"
repeat_region	20466. .20717
	/rpt_family="L2"
repeat_region	21105. .21177
	/rpt_family="L2"
repeat_region	21500. .21711
	/rpt_family="MIR"
repeat_region	21714. .22008
	/rpt_family="Alu"
repeat_region	23116. .24228
	/rpt_family="L1"
repeat_region	29132. .29495
	/rpt_family="MaLR"
misc_feature	30420. .30575
	/note="similar to EST AW426876 (NID:g6954823)"
misc_feature	31440. .31542
	/note="similar to EST AW426876 (NID:g6954823)"
repeat_region	32165. .32351
	/rpt_family="MER2_type"
repeat_region	33279. .33491
	/rpt_family="MIR"
repeat_region	34346. .34404
	/rpt_family="ERVL"
repeat_region	34528. .34641
	/rpt_family="L2"
repeat_region	34935. .35134
	/rpt_family="MER1_type"
repeat_region	35346. .35425
	/rpt_family="L2"
repeat_region	36253. .36360
	/rpt_family="MIR"
repeat_region	36392. .36473
	/rpt_family="L2"
repeat_region	37732. .38360

/rpt_family="L1"

Query Match 11.0%; Score 221; DB 8; Length 202317;
Best Local Similarity 59.3%; Pred. No. 6.6e-36;
Matches 432; Conservative 0; Mismatches 290; Indels 7; Gaps 3;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| | |||||||||  ||||||||||||||||||||||||||||||| |||||
Db      86875 TGATCCACCTGACTTGGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACTACGCCCA 86934

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
               | |   ||  ||| | |          |    | |||   | | | ||    |   |
Db      86935 GCCCTGTGATTTTTTACTGGTCTCTCTTCTCCCTAGTCTTTCTTGTCTTTTTCCATCCTC 86994

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
                |         ||           | |   ||  |      | | | |    |
Db      86995 TTTTCTCTGCTCCCTCTTCAATTCTGTTTCCTCTTTCTTCTCCCTTAATAAATATATT 87054

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |  ||         || | |    | || | |   |||  |    |
Db      87055 GAGCATTCCAAACATATTTTAAAAGACTGAGATTGTTGAATTGTTTTATAGTCTTGTCCT 87114

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              ||||   |  |||         || |  ||       | |       ||| |   |
Db      87115 TAAAATATAAATTAGTGAGTTTATTATTTATTCTCTCTATTTCTCCCAAGCAAAAGAAAA 87174

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |  |    | ||   | | ||    |||    ||     |    ||    | |   |
Db      87175 TCTCCCACAATGGCAAATAAATTAATATAAGGTATTTTAAAAATATTTTGTGTGGTTACA 87234

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAAT--GGGCCAGGCGCAGTAGCTCCTGCCT 418
              |||  || | | |||  ||||      |  ||||  ||||| || ||||| ||||  ||||
Db      87235 TAAATCATTAACATTGATAAAAGTTATATAAATCAGGGCCGGGTGCAGTGGCTCACGCCT 87294

Qy        419 GTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGA 478
              ||||||||| ||||||||||||  |||  || ||||||||||||||||||||||||| |||
Db      87295 GTAATCCCAGCACTTTGGGAGGCTGAGGCAGGTGGATCACCTGAGGTCAGGAGTTTGAGA 87354

Qy        479 CCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTG 538
              |||||||||||| |||||||||||||||||||||||||||||   |||| |||| | ||||
Db      87355 CCAGCCTGGCCAGCATGGTGAAACCCTGTCTCTACTAAAAATATAAAAGTTAGTCGGGTG 87414

Qy        539 TGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAA 598
              ||| |||||||||||||| ||||| |||||||||||| ||||| |||||||| || || |
Db      87415 TGGCGGCATGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCCTGTA 87474

Qy        599 CCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGG--CAA 653
              || |||||||||||||| ||| | | ||||   || ||||||||||||||| ||    |
Db      87475 CCCAGGAGGTGGAGGTTGAAGTGAGCTGAGATTGTGCCACTGCACTCCAGCCAGGGTGAC 87534

Qy        654 CAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTA 713
               |||||  |||||||||    | ||  ||| |    || || | ||||| | ||     ||
Db      87535 AAGAGCGAGACTCCATCTCAAAAAACAAAACAAAAAAGTTATATAAATCAATTAACATTA 87594

Qy        714 TCTTCATTT 722
              |||||||
Db      87595 ATTTCATTT 87603
```

RESULT 28
AC036222
LOCUS AC036222 174965 bp DNA linear PRI 19-OCT-2002
DEFINITION Homo sapiens chromosome 17, clone RP11-433M22, complete sequence.
ACCESSION AC036222
VERSION AC036222.23 GI:24137556
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 174965)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-433M22
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174965)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 174965)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 174965)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Oct 19, 2002 this sequence version replaced gi:23957647.
All repeats were identified using RepeatMasker:

```
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ---------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          ---------- Project Information
              Center project name: L9421
              Center clone name: 433_M_22
          ----------

          Only the last 175.0 kilobases of this clone are being submitted.
          The remainder overlaps accession number AC103702 [WICGR project
          L21590].
FEATURES             Location/Qualifiers
     source          1. .174965
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-433M22"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   8. .136
                     /rpt_family="L2"
     repeat_region   complement(436. .549)
                     /rpt_family="MIR"
     repeat_region   570. .859
                     /rpt_family="AluSx"
     repeat_region   860. .888
                     /rpt_family="AT_rich"
     repeat_region   complement(1035. .1254)
                     /rpt_family="L1MC4"
     repeat_region   complement(1299. .1970)
                     /rpt_family="L1MB2"
     repeat_region   complement(1971. .2266)
                     /rpt_family="AluSx"
     repeat_region   2267. .2325
                     /rpt_family="(TTG)n"
     repeat_region   complement(2328. .2608)
                     /rpt_family="AluSq"
     repeat_region   complement(2609. .2932)
                     /rpt_family="L1MB2"
     repeat_region   complement(2943. .3230)
                     /rpt_family="AluSc"
     repeat_region   3287. .3560
                     /rpt_family="L2"
     repeat_region   complement(3704. .3857)
                     /rpt_family="MIR"
     repeat_region   3858. .4159
                     /rpt_family="AluSp"
     repeat_region   complement(4160. .4205)
                     /rpt_family="MIR"
     repeat_region   4206. .4571
                     /rpt_family="L3"
     repeat_region   4677. .4965
                     /rpt_family="AluJo"
     repeat_region   complement(6458. .6602)
                     /rpt_family="MIR"
     repeat_region   6675. .6778
                     /rpt_family="L1MC5"
     repeat_region   complement(6817. .6859)
                     /rpt_family="L3"
     repeat_region   complement(6910. .7127)
                     /rpt_family="MIR"
     repeat_region   complement(7141. .7437)
                     /rpt_family="AluSx"
     repeat_region   complement(8922. .9145)
                     /rpt_family="MIR"
     repeat_region   9466. .9595
                     /rpt_family="FLAM_C"
     repeat_region   10532. .10633
                     /rpt_family="MIR"
     repeat_region   12389. .12450
                     /rpt_family="MER5A"
     repeat_region   12451. .12761
                     /rpt_family="AluJo"
     repeat_region   12762. .12860
                     /rpt_family="MER5A"
     repeat_region   complement(12869. .12914)
                     /rpt_family="MIR"
     repeat_region   complement(12915. .13156)
                     /rpt_family="AluSx"
     repeat_region   complement(13157. .13310)
                     /rpt_family="MIR"
     repeat_region   13546. .13604
                     /rpt_family="L2"
     repeat_region   14682. .14725
                     /rpt_family="(TATAA)n"
     repeat_region   14929. .15043
                     /rpt_family="AluJb"
     repeat_region   15044. .15074
                     /rpt_family="AT_rich"
     repeat_region   15463. .15484
                     /rpt_family="L2"
     repeat_region   complement(15485. .15787)
                     /rpt_family="AluY"
     repeat_region   15788. .15869
                     /rpt_family="L2"
     repeat_region   complement(16233. .16368)
                     /rpt_family="L1MA10"
     repeat_region   complement(16384. .16538)
                     /rpt_family="MIR3"
     repeat_region   16644. .16697
                     /rpt_family="CT-rich"
     repeat_region   complement(16771. .17063)
                     /rpt_family="AluSq"
     repeat_region   complement(17066. .17263)
                     /rpt_family="MIR"
     repeat_region   17267. .17293
                     /rpt_family="(TTTG)n"
     repeat_region   17425. .17562
                     /rpt_family="L2"
     repeat_region   complement(18106. .18304)

  Query Match           11.0%;  Score 220.6;  DB 8;  Length 174965;
  Best Local Similarity  61.2%;  Pred. No. 8.2e-36;
  Matches  430;  Conservative    0;  Mismatches  259;  Indels   14;  Gaps    4;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| ||||| |||| ||||||||||||||||||||||||||||| |||  |||| |
Db      68046 TGATCCACCCGCCTTAGCCTTCCAAAGTGCTGGGATTACAGGCGTGAGTCACTGCGCCCG 68105

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  |       |  ||    ||  || ||  | ||          | |  ||
Db      68106 GCCTCGTACTAGGTGTTATTTTTAAGGGAATATCCTCATTCTTAGAAAATACCGTT--GA 68163

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||  |||   | |     |    |   |||  | |||||          ||  |  |
Db      68164 AGTTTTTAGGGAAAAGGAGCGTGATGACTGTAATTTACTCTCAAATTCAGAAAAAAATGT 68223

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |          ||   | | | | ||   |  |    | ||   ||   |  |  ||
Db      68224 AAATTAATCTAGTACTAATTACTTTTCTATAATTAACATTTTATTTATTATGAAAAATAA 68283

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |  | | | || || |||  ||    ||| |||   |||||          | | | |
Db      68284 AGTAAATTTAATAATTGATGATTATTAAATTTAATAATTAATAATCGAAGTTAATAATTG 68343

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTA--------TTTAATAAAATT 352
              | | | |    |    ||    | | || |    ||||          |||    || ||
Db      68344 ATGATTCCAAGTGAATGGTATACTTGAGTTCTAGATACTATGATAACTTTTCTGTAAGTT 68403

Qy        353 TTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTC 412
              | |||   |   | | | |||||||     |        ||  ||||||||   || ||||
```

```
Db      68404 TGAAATTATTTCAAAATAAAATTATTTTTAAAAAACCCAACAGGCCAGGCATGGTGGCTC 68463

Qy        413 CTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGT 472
              |||||||||||||||| |||||||||||||||| | ||||||| ||||||||||||
Db      68464 ACGCCTGTAATCCCAACATTTTGGGAGGCCAAGGTGAGTGGATCACTTGAGGTCAGGAGT 68523

Qy        473 TTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGC 532
              | |||| |||||||||||||||| |||||||||||||||||||||  ||||||||||
Db      68524 TCGAGACAAGCCTGGCCAACATGGAGAAACCCTGTCTCTACTAAAAATACAAAAATTAGC 68583

Qy        533 CAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTG 592
              |||||||||||||| ||||||||| || || ||| || | ||  ||||| ||||||||||
Db      68584 CAGGTGTGGTGGCACGCACCTGTAATCTCAGCTATTCGGAAGTCTGAGGCAGGAGAATTG 68643

Qy        593 CTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCAC---TGCACTCCAGCCT- 648
              |||||||| ||||||  |||||| |||| | |||||||  |||   ||||||||||||
Db      68644 CTTGAACCCAGGAGGCAGAGGTTACAGTGAGCCGAGATCGCACCATTGCACTCCAGCCTG 68703

Qy        649 GGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAA 691
              ||  ||||||||||||||  |    |  | | ||| |    |||
Db      68704 GGTGACAGAGCAAGACTCTGTCTCAAACAAACAAACAAACAAA 68746
```

```
RESULT 29
AC004803/c
LOCUS       AC004803              154369 bp    DNA     linear   PRI 25-FEB-2003
DEFINITION  Homo sapiens 12 BAC RP11-359B12 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC004803
VERSION     AC004803.11  GI:28557808
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 154369)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
            Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
            Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
            Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
            Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
            Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
            King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
            Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
            Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X.,
            Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
            Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
            Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
            Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
            Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
            Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154369)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 154369)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 154369)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 25, 2003 this sequence version replaced gi:23957598.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
            ml.
FEATURES             Location/Qualifiers
     source          1. .154369
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-359B12"
     misc_feature    1. .2210
                     /note="overlaps bases 148591. .150804 of clone AC004765"
                     /function="clone overlap"
```

```
repeat_region   357. .377
                /rpt_family="(T)n"
STS             1905. .2020
                /standard_name="G41230"
repeat_region   complement(3042. .3318)
                /rpt_family="AluSx"
STS             3795. .4057
                /standard_name="G41780"
repeat_region   4718. .4874
                /rpt_family="AluSg/x"
repeat_region   10377. .10663
                /rpt_family="AluY"
repeat_region   complement(10792. .11092)
                /rpt_family="AluSx"
repeat_region   complement(11984. .12277)
                /rpt_family="AluSp"
repeat_region   12434. .12467
                /rpt_family="(T)n"
repeat_region   12537. .12610
                /rpt_family="Charlie1"
repeat_region   12613. .12777
                /rpt_family="AluJo"
repeat_region   12778. .12879
                /rpt_family="Charlie1"
repeat_region   complement(12924. .13070)
                /rpt_family="L1PA4"
repeat_region   complement(13098. .13292)
                /rpt_family="MER58A"
repeat_region   13813. .14115
                /rpt_family="AluJb"
repeat_region   14120. .14220
                /rpt_family="L1PA15"
repeat_region   14382. .14663
                /rpt_family="AluY"
repeat_region   14664. .14687
                /rpt_family="(TAA)n"
repeat_region   14776. .15068
                /rpt_family="AluSx"
repeat_region   16255. .16286
                /rpt_family="AT_rich"
repeat_region   complement(16287. .16598)
                /rpt_family="AluSx"
STS             17389. .17673
                /standard_name="D12S888"
repeat_region   17727. .17779
                /rpt_family="(CA)n"
repeat_region   17863. .17943
                /rpt_family="(TTA)n"
repeat_region   19043. .19345
                /rpt_family="AluY"
repeat_region   complement(19413. .19584)
                /rpt_family="AluJo/FRAM"
repeat_region   complement(19588. .19879)
                /rpt_family="AluSx"
repeat_region   complement(20140. .20420)
                /rpt_family="AluSx"
repeat_region   20849. .20875
                /rpt_family="(T)n"
repeat_region   complement(20876. .21127)
                /rpt_family="AluSq"
repeat_region   22856. .23154
                /rpt_family="AluJb"
STS             23557. .23689
                /standard_name="STS-N22269"
repeat_region   23646. .23685
                /rpt_family="L2"
repeat_region   complement(25629. .25725)
                /rpt_family="L2"
STS             26672. .26904
                /standard_name="G41789"
repeat_region   27598. .27877
                /rpt_family="AluJo"
repeat_region   27878. .28175
                /rpt_family="AluSx"
STS             30213. .30390
                /standard_name="D12S1913"
repeat_region   31536. .31559
                /rpt_family="(CCA)n"
STS             31816. .31981
```

Query Match 11.0%; Score 220; DB 8; Length 154369;
Best Local Similarity 60.0%; Pred. No. 1.1e-35;
Matches 420; Conservative 0; Mismatches 275; Indels 5; Gaps 3;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||| ||||||||||||||||||| || |||| ||||||||  | ||
Db      78078 TGATCCACCTGCCTCGGCCTCCCAAAGTGCTGGGAATATAGGCATGAGCCACTGCACCCA 78019

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  |       || |  || |  ||  ||.||      ||   ||| | |  |||||| |
Db      78018 GCCTCAAAATGGTAATTCTTAATAAGTGAAAATTTTATTTCCTTCT-TTTCTTATTTGGC 77960

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||  |      ||       |      |  |   || | ||      | | |
Db      77959 TTTTTTAAACAACTGAAATCTGAAGGTAGTTGGCCCAGATAAGGACTTTTGTCTGCATGG 77900

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | | |  |||    | ||          ||    ||  | | |  |   |  ||  ||
Db      77899 AATTTTTATGCTTCTCTCCTTTTATCAGGTGGGTTCATTGTAAGGTGTGGAAGTGAAAAA 77840

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              ||  | ||  ||  | |   ||       | ||  || | | |||  |  | |   | |
Db      77839 GAATGCAATTAACATTACAGAAACATCTATTAAATGTCAGATATTGTGTGCTTTATATCT 77780

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              | ||  ||   |  || | | |  |  |     |  |     | |  || |   | |
Db      77779 CATCTAATAACTATAATCTTATAACTCATCTAATCTTCCCACAACTCTAACTAGTAGATG 77720

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              | ||  | |   | | ||||  |  |   ||  |||||||||||| || ||||  ||||||
Db      77719 TTATTTATTCTCACT-TTAAGAATATGGAAATGGGGCCAGGCGCGGTGGCTCACGCCTGT 77661

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||| |||| ||| ||| |||||||||||||||| ||||||
Db      77660 AATCCCAGCACTTTGGGAGGCCGAGGTAGGCAGATAACCTGAGGTCAGGAGTTCAAGACC 77601

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db      77600 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGTGTG 77541

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              |||||  |  ||| || ||||| |||||  ||||| |||| ||||  || |||||| ||
Db      77540 GTGGCGGGTGCCTATAATCCCAGCTACTTGGGAGGCTGAGGCAGGAAGATCGCTTGAGCC 77481

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGA 657
              ||||  |||||||||||| | || | ||     ||||||||||||||||||     ||
Db      77480 CGGGAGACGGAGGTTGCAGTGAGCCAATATCACACCACTGCACTCCAGCCTGAGTCAAGC 77421

Qy        658 GCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGT 697
                |  | |      | |||||||      | | |||||
Db      77420 AGCAAGATTCTGTCTCAAAACAAAACAAAACAGAATGTGT 77381
```

```
RESULT 30
AC092035
LOCUS       AC092035              162997 bp    DNA     linear   PRI 07-DEC-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-72H11, complete sequence.
ACCESSION   AC092035 AC023893
VERSION     AC092035.2  GI:17402776
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 162997)
```

AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162997)
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 162997)
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT On Dec 7, 2001 this sequence version replaced gi:14475968.
---------- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchtgs@u.washington.edu
Drafting Center: BCM
---------- Project Information
Center project name: chr-3
Center clone name: RP11-72H11 (bc0191)
---------- Summary Statistics
Sequencing vector: M13; L08821; 49% of reads
Sequencing vector: plasmid; L08752; 51% of reads
Chemistry: Dye-terminator ET; 31% of reads
Chemistry: Dye-terminator Big Dye; 69% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 162937 bases at least Q40
Consensus quality: 162993 bases at least Q30
Consensus quality: 162997 bases at least Q20
Insert size: 189816; 14.3% error; agarose-fp
Insert size: 170535; sum-of-contigs
Quality coverage: 8.5x in Q20 bases; agarose-fp
Quality coverage: 9.5x in Q20 bases; sum-of-contigs
----------.
Overlapping Sequences:
5': Mapping in progress
3': RP11-782N4 (UWGC:bc0561) AC023231

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI		HindIII		BglII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
8696	8877	9015	9409	17217	16762
6	<800	6382	6539	2067	2096
2241	2268	512	<800	4800	5125
9160	8877	449	<800	392	<800
627	<800	7149	7144	5065	4984
12095	11762	10802	10131	2902	2898
4269	4232	172	<800	5106	5125
3622	3617	1312	1235	399	<800
76	<800	299	<800	5472	5427
8290	8877	574	<800	1257	1274
3310	3248	539	<800	15777	16762
1177	1194	5499	5445	913	876
471	<800	2632	2668	1851	1888
3341	3248	505	<800	302	<800
3806	3744	1570	1621	5971	5922
8735	8877	1662	1748	2385	2351
9467	9473	1925	1948	8553	8469
4680	4647	287	<800	129	<800
1777	1789	5559	5445	2327	2351
13241	12854	404	<800	4257	4244
12983	12854	974	891	1128	1139
2918	2928	1457	1446	588	<800
17847	18077	349	<800	2088	2096
2843	2802	741	<800	4796	4984
10470	10353	50	<800	3063	3069
1195	1194	2476	2519	281	<800
1940	1970	567	<800	2172	2223
4982	4986	543	<800	93	<800
8902	8877	1890	1948	8689	8469
8529	8877	4417	4370	1940	1982
		1391	1446	9568	9198
		10779	10131	755	<800
		1461	1446	3576	3554
		1033	975	1018	926

1823	1948	2555	2605
3444	3493	117	<800
3859	3917	2023	2096
19991	20261	7990	8018
3008	3067	1044	1050
8698	8729	4448	4427
378	<800	1507	1529
4338	4370	8086	8018
8943	8729	562	<800
1891	1948	1312	1274
2502	2519	15155	14844
408	<800		
1360	1446		
3017	3067		
2486	2519		
9784	10131		
1080	1027		
438	<800		
990	975		
1196	1122		
215	<800		

Query Match 11.0%; Score 220; DB 8; Length 162997;
Best Local Similarity 80.4%; Pred. No. 1.1e-35;
Matches 283; Conservative 0; Mismatches 65; Indels 4; Gaps 2;

```
Qy      389 GAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTG 448
            ||  |||||||||| || | ||||  |||||||||||| |||||||||||||| ||| |
Db    37689 GATGTGGGCCAGGCACAATGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCG 37748

Qy      449 GGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTC 508
            ||| ||||||||||||||||||||| |||||||||||| ||||||||||||||||  ||
Db    37749 GGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGGCAACATGGTGAAACCCCATC 37808

Qy      509 TCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACT 568
            ||||||||||  |||||| ||||||||||||||||||  ||||||| | ||||  |||||
Db    37809 TCTACTAAAAATACAAAAACTAGCCAGGTGTGGTGGCCGGCACCTGCAATCCCTGCTACT 37868

Qy      569 CAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG 628
            |||||||||||| |||||||||||||||||| || |||||||||||||| | | |||
Db    37869 CAGGAGGTTGAGGCAGGAGAATTGCTTGAACCTGGGGGGTGGAGGTTGCAGTGAGCTGAG 37928

Qy      629 ---ATGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAAGC 684
               ||  |||||||||||||| | |||||||||||||||||||| || | || ||||
Db    37929 ATCATACCACTGCACTCCAGCTTGGGCAACAGAGCAAGACTCCATTAAAAAAAAAAAAAA 37988

Qy      685 TTTGAAATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGAAATTCATCTT 736
                |||      |||  || |||    || ||  | | | | ||| || ||
Db    37989 AAAAAAAAAAAAAAAACAAGATGTGAAGATTTTTCTATCATATATTAATGTT 38040
```

RESULT 31
AC067852/c
LOCUS AC067852 179146 bp DNA linear PRI 27-FEB-2003
DEFINITION Homo sapiens chromosome 17, clone RP11-400F19, complete sequence.
ACCESSION AC067852
VERSION AC067852.23 GI:28570403
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179146)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-400F19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179146)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 179146)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 179146)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,

```
          Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
          Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
          Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
          Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
          Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
          O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
          Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
          Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
          Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
          Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
          Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
          Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (27-FEB-2003) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Feb 27, 2003 this sequence version replaced gi:28173140.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L9335
              Center clone name: 400_F_19
          --------------.
FEATURES             Location/Qualifiers
     source          1. .179146
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-400F19"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   138. .440
                     /rpt_family="AluJo"
     repeat_region   962. .1001
                     /rpt_family="GC_rich"
     repeat_region   1058. .1081
                     /rpt_family="GC_rich"
     repeat_region   1327. .1382
                     /rpt_family="CT-rich"
     repeat_region   2432. .2469
                     /rpt_family="GC_rich"
     repeat_region   2902. .2934
                     /rpt_family="(AGGGGG)n"
     repeat_region   complement(3214. .3365)
                     /rpt_family="MIR3"
     repeat_region   complement(5216. .5317)
                     /rpt_family="MIR"
     repeat_region   5700. .5721
                     /rpt_family="GC_rich"
     repeat_region   complement(6114. .6215)
                     /rpt_family="MER91A"
     repeat_region   7761. .7848
                     /rpt_family="(CCCCG)n"
     unsure          7868. .7873
                     /note="<30 qual SNGL region"
     unsure          7874. .7931
                     /note="<30 qual single clone coverage"
     repeat_region   7893. .7930
                     /rpt_family="GC_rich"
     repeat_region   complement(8646. .8917)
                     /rpt_family="L2"
     repeat_region   8918. .9216
                     /rpt_family="AluSc"
     repeat_region   complement(9217. .9271)
                     /rpt_family="L2"
     repeat_region   9272. .9569
                     /rpt_family="AluY"
     repeat_region   complement(9570. .9622)
                     /rpt_family="L2"
     repeat_region   12682. .13009
                     /rpt_family="AluY"
     repeat_region   13015. .13134
                     /rpt_family="FLAM_C"
     repeat_region   13143. .13185
                     /rpt_family="(TA)n"
     repeat_region   complement(13251. .13398)
                     /rpt_family="L2"
     repeat_region   complement(13736. .14068)
                     /rpt_family="AluJb"
     repeat_region   complement(14071. .14366)
                     /rpt_family="AluSx"
     repeat_region   complement(14367. .14502)
                     /rpt_family="AluJb"
     repeat_region   complement(15104. .15222)
                     /rpt_family="MER21B"
     repeat_region   15223. .15277
                     /rpt_family="Tigger5"
     repeat_region   15247. .15319
                     /rpt_family="Tigger5"
     repeat_region   complement(15409. .15704)
                     /rpt_family="AluSx"
     repeat_region   complement(15710. .16000)
                     /rpt_family="AluJo"
     repeat_region   complement(16010. .16066)
                     /rpt_family="Tigger5"
     repeat_region   complement(16069. .16668)
                     /rpt_family="MER21B"
     repeat_region   18847. .18889
                     /rpt_family="G-rich"
     repeat_region   complement(18986. .19008)
                     /rpt_family="L1MB3"
     repeat_region   complement(19009. .19048)
                     /rpt_family="L1MB3"
     repeat_region   complement(19049. .19352)
                     /rpt_family="AluJo"
     repeat_region   complement(19353. .19508)
                     /rpt_family="L1MB3"
     repeat_region   complement(19509. .19929)
                     /rpt_family="L1MB3"
     repeat_region   19931. .20323
                     /rpt_family="LTR54"
     repeat_region   complement(20327. .20627)
                     /rpt_family="AluSx"
     repeat_region   complement(20645. .20929)
                     /rpt_family="AluSp"
     repeat_region   complement(20930. .21218)
                     /rpt_family="AluSg"
     repeat_region   complement(21233. .21322)
                     /rpt_family="AluJo/F(R)AM"
     repeat_region   complement(21323. .21614)
                     /rpt_family="AluSx"
     repeat_region   complement(21615. .22365)

  Query Match            11.0%;  Score 220;  DB 8;  Length 179146;
  Best Local Similarity  60.8%;  Pred. No. 1.1e-35;
  Matches  435;  Conservative    0;  Mismatches  265;  Indels   16;  Gaps    4;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||| || ||||  ||||||||||||||||||||||||||||||||||||||| | || |
Db      43622 TGATCCGCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCACCCG 43563

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | | |  || |   | |||| |  |  |||           || ||    ||   | |
Db      43562 GCCTACAGTGT---TGATTTAACAAACCAAATGATAGTAGGAAACTAAAATCATGTCAAA 43506

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
               |   || |  |   |||  |     |     ||| |  |     ||   |  |  |
Db      43505 TGTTAATGCTTAATAACAATATCCTAAATAGTTTTTAATTTAAAAATTTGCCTAGCACTG 43446
```

```
Qy        181 TCTCTCTCAGCTGTCTTTCTGCATAT--------GTAAGGTGGAATAATGATGGCTTTCA 232
              | |  || | ||||  | || | ||          |   |  | || |     ||  |
Db      43445 TGTTACTAACCTGTAGTCCTACCTACTACTAGGAGGCTGAGGCAAGAGAATTGCTTCAGC 43386

Qy        233 AGGAGGAATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCA 292
                |||| |  |    || |  | | |  |   |        | |    |    |
Db      43385 CCCAGGAGTTCAAGGTTGCAGTGAGCTATGATCATGCCACTACACTCCAGTCTGGGCAAC 43326

Qy        293 ACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATT 352
              | |  | |  |        || ||  |  | | | | || |  |  | ||||   ||
Db      43325 AGAGAGACTCTGTTTCTACAAAAATAAATAAATAAGTAATAAATATCAATTAAATCAAAA 43266

Qy        353 TTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTC 412
              |  |  | |       | |    |||| |   | |||   |||||||| ||| ||||
Db      43265 TGGACTAAAATTCTCAACATACATATAAAAATAATGGAAGCAGGCCAGGCGTAGTGGCTC 43206

Qy        413 CTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGT 472
              |||||||||||| |||||||||||||||| |||||||||||| ||||||||||
Db      43205 ACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGATGGGCGGATCACCTAAGGTCAGGAGT 43146

Qy        473 TTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGC 532
              | ||||||||||||||||||||  |||||||||||||||||||  ||||||||| |
Db      43145 TCAAGACCAGCCTGGCCAACATGGCAAAACCCTGTCTCTACTAAAAATACAAAAATTATC 43086

Qy        533 CAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTG 592
              |||||||||||  |  |||||| ||||| |||||| ||||| ||||| ||||||||  |
Db      43085 TGGGTGTGGTGGCAGGTGCCTGTAATCCCACCTACTCTGGAGGCTGAGGCAGGAGAATCG 43026

Qy        593 CTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCAC---TGCACTCCAGCCTG 649
              |||||||| ||||| |||||||| ||| | | |||||  |||   ||||||||||||||
Db      43025 CTTGAACCTGGGAGGAGGAGGTTGTAGTGAGCTGAGATTGCACCATTGCACTCCAGCCTG 42966

Qy        650 G--CAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGA 703
              |   |  ||||||||||||||||    | ||  ||   |  ||| |  |||  ||
Db      42965 GGCGACAAGAGCAAGACTCCATCTCAAAAAATAATAATAATAAAATAAATAATGGA 42910
```

```
RESULT 32
BV651164/c
LOCUS       BV651164                841 bp    DNA     linear   STS 16-APR-2005
DEFINITION  S217P61861RA5.T0 Noemie Pan troglodytes troglodytes STS genomic,
            sequence tagged site.
ACCESSION   BV651164
VERSION     BV651164.1  GI:62679133
KEYWORDS    STS.
SOURCE      Pan troglodytes troglodytes
  ORGANISM  Pan troglodytes troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 841
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES             Location/Qualifiers
     source          1. .841
                     /organism="Pan troglodytes troglodytes"
                     /mol_type="genomic DNA"
                     /sub_species="troglodytes"
                     /db_xref="taxon:37011"
                     /clone_lib="Noemie"
     STS             <1. .>841
ORIGIN
```

```
  Query Match            11.0%;  Score 219.2;  DB 10;  Length 841;
  Best Local Similarity  70.1%;  Pred. No. 4.4e-35;
  Matches  324;  Conservative    0;  Mismatches  133;  Indels    5;  Gaps    2;

Qy        215 GAATAATGATGGCTTTCAAGGAGGAATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGA 274
              ||  ||   |   | ||||  |   | |||   || |||| || | || | |      ||
Db        464 GAGCAAGACTCCGTCTCAAAAAAAGAAAAATAAATCAAAAATGGTTAGAAGAAATAAAGA 405

Qy        275 TATGAAATAAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATTTAAGAGT--TATTTAT 332
                  | ||   |||   | ||||   |          |||||| ||| | |  ||  |||
Db        404 AGGACATTATATATTAATAAAAGTTTCAATATAAGAAGAAGATATAACAATTGTAAATAT 345

Qy        333 TACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAA 392
              |   ||   | |  | |     ||| ||    |   |||  ||| |      |  ||
Db        344 TTACACACCTAATGACAGACCATCAAAATATCTGAAGCAAAAAATTGATAGAACTGAAAG 285

Qy        393 TGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCG 452
              |||||||||||||| |||| || |||||||| || ||||||||||||||||||||||||
Db        284 AAGGCCAGGCGCAGTGGCTCATGACTGTAATCTCAGCACTTTGGGAGGCCAAGGTGGGCA 225

Qy        453 GATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTA 512
              ||||||  |||||||||||||| |||||||||||||||||| |||  ||||||||||||||
Db        224 GATCACTTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACGTGGCAAAACCCTGTCTCTA 165

Qy        513 CTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGG 572
              |||||||  ||||||||||||||  ||||||| ||  | |||| |||||  |||||||||
Db        164 CTAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGTGCTTGTAATCCCAGCTACTCAGG 105

Qy        573 AGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA--- 629
              ||  ||||  |||||||| || ||||||  |||||||| |||||||||| | | ||||
Db        104 AGACTGAGTCAGGAGAATCGCCTGAACCCAGGAGGTGAAGGTTGCAGTGAACAGAGATCG 45

Qy        630 TGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAA 671
              || ||||||||||||||||||||||||||||||||||||||| ||
Db         44 TGCCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCAGAA 3
```

```
RESULT 33
AL359713
LOCUS       AL359713              152966 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-95P3 on chromosome 6 Contains
            the 5' end of the RU2 gene, the MRS2L gene for MRS2-like magensium
            homeostasis factor, the 3' end of the GPLD1 gene for
            glycosylphosphatidylinositol specific phospholipase D1 and two CpG
            islands, complete sequence.
ACCESSION   AL359713
VERSION     AL359713.25  GI:13938809
KEYWORDS    HTG; CpG island; glycosylphosphatidylinositol specific
            phospholipas; GPIPLD; GPLD1; HPT; magnesium homeostasis factor;
            MRS2L; phosphatidylinositol-glycan-specific phospholipase; PIGPLD;
            PIGPLD1; RU2; RU2AS; RU2S.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 152966)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 3, 2001 this sequence version replaced gi:13446455.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-95P3 is from the library RPCI-11.1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .152966
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-95P3"
                     /clone_lib="RPCI-11.1"
     misc_feature    1
                     /note="Clone_left_end: RP11-95P3"
     misc_feature    17700
                     /note="Clone_right_end: RP11-40E20"
     polyA_site      complement(49750)
     gene            complement(join(49751. .49838,50088. .50142,53977. .>54172))
                     /gene="DCDC2"
                     /locus_tag="RP11-260H5__B.1-003"
     mRNA            complement(join(49751. .49838,50088. .50142,53977. .>54172))
                     /gene="DCDC2"
                     /locus_tag="RP11-260H5__B.1-003"
                     /product="doublecortin domain containing 2"
                     /note="match: ESTs: AW293816"
     polyA_signal    complement(49765. .49770)
                     /gene="DCDC2"
                     /locus_tag="RP11-260H5__B.1-003"
     CDS             complement(join(49824. .49838,50088. .50142,53977. .>54172))
                     /gene="DCDC2"
                     /locus_tag="RP11-260H5__B.1-003"
                     /standard_name="OTTHUMP00000017889"
                     /codon_start=3
                     /product="doublecortin domain containing 2"
                     /protein_id="CAI17105.1"
                     /db_xref="GI:55959678"
                     /db_xref="GOA:Q5T3Y5"
                     /db_xref="InterPro:IPR003533"
                     /db_xref="UniProt/TrEMBL:Q5T3Y5"
                     /translation="VVIHEKKVSSFEVFLKEVTGGVQAPFGAVRNIYTPRTGHRIRKL
                     DQIQSGGNYVAGGQEAFKKLNYLDIGEIKKRPMEVVNTELHVL"
     gene            join(complement(53977. .54571),complement(50088. .50142),
                     complement(AL359389.15:80898. .80974),
                     complement(AL359389.15:80645. .80776),
                     complement(AL359389.15:69862. .70008),
                     complement(AL359389.15:67782. .67836),
                     complement(AL359389.15:56979. .57141),
                     complement(AL591344.1:45608. .45708),
                     complement(AL591344.1:18936. .19238),
                     complement(AL591344.1:12362. .15440))
                     /gene="DCDC2"
                     /locus_tag="RP11-260H5__B.1-001"
     mRNA            join(complement(53977. .54571),complement(50088. .50142),
                     complement(AL359389.15:80898. .80974),
                     complement(AL359389.15:80645. .80776),
                     complement(AL359389.15:69862. .70008),
                     complement(AL359389.15:67782. .67836),
                     complement(AL359389.15:56979. .57141),
                     complement(AL591344.1:45608. .45708),
                     complement(AL591344.1:18936. .19238),
                     complement(AL591344.1:12362. .15440))
                     /gene="DCDC2"
                     /locus_tag="RP11-260H5__B.1-001"
                     /product="doublecortin domain containing 2"
                     /note="match: ESTs: AA004587 AA004717 AI889161 AL564319
                     AV750645 BF593580 BF876266 BI760196 BM468677
                     match: cDNAs: AF181721.1 AK027036.1 BC014954.1"
     CDS             join(complement(53977. .54269),complement(50088. .50142),
                     complement(AL359389.15:80898. .80974),
                     complement(AL359389.15:80645. .80776),
                     complement(AL359389.15:69862. .70008),
                     complement(AL359389.15:67782. .67836),
                     complement(AL359389.15:56979. .57141),
                     complement(AL591344.1:45608. .45708),
                     complement(AL591344.1:18936. .19238),
                     complement(AL591344.1:15336. .15440))
                     /gene="DCDC2"
                     /locus_tag="RP11-260H5__B.1-001"
                     /standard_name="OTTHUMP00000017887"
                     /note="match: proteins: Q9NU80 Q9UHG1 Q9ULR6"
                     /codon_start=1
                     /product="doublecortin domain containing 2"
                     /protein_id="CAI17104.1"
                     /db_xref="GI:55959677"
                     /db_xref="GOA:Q9UHG0"
                     /db_xref="UniProt/Swiss-Prot:Q9UHG0"
                     /translation="MSGSSARSSHLSQPVVKSVLVYRNGDPFYAGRRVVIHEKKVSSF
                     EVFLKEVTGGVQAPFGAVRNIYTPRTGHRIRKLDQIQSGGNYVAGGQEAFKKLNYLDI
                     GEIKKRPMEVVNTEVKPVIHSRINVSARFRKPLQEPCTIFLIANGDLINPASRLLIPR
                     KTLNQWDHVLQMVTEKITLRSGAVHRLYTLEGKLVESGAELENGQFYVAVGRDKFKKL
                     PYSELLFDKSTMRRPFGQKASSLPPIVGSRKSKGSGNDRHSKSTVGSSDNSSPQPLKR
                     KGKKEDVNSEKLTKLKQNVKLKNSQETIPNSDEGIFKAGAERSETRGAAEVQEDEDTQ
                     VEVPVDQRPAEIVDEEEDGEKANKDAEQKEDFSGMNGDLEEEGGREATDAPEQVEEIL
                     DHSEQQARPARVNGGTDEENGEELQQVNNELQLVLDKERKSQGAGSGQDEADVDPQRP
                     PRPEVKITSPEENENNQQNKDYAAVA"
     gene            53422. .54803
                     /gene="KAAG1"
```

```
                     /locus_tag="RP11-95P3.2-001"
     mRNA            53422. .54803
                     /gene="KAAG1"
                     /locus_tag="RP11-95P3.2-001"
                     /product="kidney associated antigen 1"
                     /note="match: cDNAs: AF181722.1"
     CDS             54159. .54413
                     /gene="KAAG1"
                     /locus_tag="RP11-95P3.2-001"
                     /standard_name="OTTHUMP00000016084"
                     /note="match: proteins: Q9UBP8"
                     /codon_start=1
                     /product="kidney associated antigen 1"
                     /protein_id="CAI17106.1"
                     /db_xref="GI:55959679"
                     /db_xref="Genew:21031"
                     /db_xref="GOA:Q9UBP8"
                     /db_xref="UniProt/TrEMBL:Q9UBP8"
                     /translation="MDDDAAPRVEGVPVAVHKHALHDGLRQVAGPGAAAAHLPRWPPP
                     QLAASRREAPPLSQRPHRTQGAGSPPETNEKLTNPQVKEK"
     polyA_signal    54782. .54787
                     /gene="KAAG1"
                     /locus_tag="RP11-95P3.2-001"
     polyA_site      54804
     misc_feature    55358. .55951
                     /note="Tandem repeat. Sequence from overlapping clone
                     dJ282H10 (AL132672). Assembly confirmed by restriction
                     digest."
     gene            join(99473. .99755,101687. .101760,104927. .104963,
                     105980. .106092,108741. .108914,111552. .111682,
                     112916. .113032,114603. .114755,114980. .115097,
                     119456. .119569,120103. .122101)
                     /locus_tag="RP11-95P3.1-009"
     mRNA            join(99473. .99755,101687. .101760,104927. .104963,
                     105980. .106092,108741. .108914,111552. .111682,
                     112916. .113032,114603. .114755,114980. .115097,
                     119456. .119569,120103. .122101)
                     /locus_tag="RP11-95P3.1-009"
                     /note="match: ESTs: AA280673 AA781573 AA782213 AA890448
                     AA933950 AA994525 AI027389 AI342569 AI522078 AI525690
                     AI628970 AI954726 AV712957 AW131633 AW295997 AW632241
                     AW974621 BE168047 BE219330 BE463602 BE798677 BF116203
                     BF216698 BF218387 BF245307 BF245875 BG421926 BG740291
                     BM018867
                     match: cDNAs: AF288288.1 AF288289.1 AF288290.1 AF288291.1
                     AF293076.1"
     gene            join(99473. .99755,101687. .101760,104927. .104963,
                     105980. .106092,108741. .108914,111552. .111682,
                     112916. .113032,114603. .114755,114980. .115097,
                     119456. .119995)
                     /locus_tag="RP11-95P3.1-011"
     mRNA            join(99473. .99755,101687. .101760,104927. .104963,
                     105980. .106092,108741. .108914,111552. .111682,
                     112916. .113032,114603. .114755,114980. .115097,
                     119456. .119995)
                     /locus_tag="RP11-95P3.1-011"
                     /note="match: ESTs: AI188501 AI206137 AI307274 AI624310
```

```
  Query Match           11.0%;  Score 219.2;  DB 8;  Length 152966;
  Best Local Similarity   59.8%;  Pred. No. 1.6e-35;
  Matches  427;  Conservative    0;  Mismatches  273;  Indels   14;  Gaps    3;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||||||  |||  ||||||||||||||||||||||||||| ||||||||||||||
Db     130251 TGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGCCCA 130310

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  || |||  |||  ||||  || |             |   |  |   |      |
Db     130311 GCCTCATTTCTACATTTCTTGATTAGCCTTACAAATGTGCATGCCCTTTGACCCCAGAAA 130370

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
                | || ||   | |   ||        ||   || |     |   | | |
Db     130371 TTCTACTTATAGAAATACAGGTCAGAAAAATAACCAGATGTATAAAAACATTTATGTGAC 130430
```

```
Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | | |  ||  |   ||    ||      ||   | |||              ||  | |.
Db     130431 TATTTACCATATTATTTAAAACAATGAAAATTGGAAATGGCACAATTGACTAACAATGGT 130490

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |      || |   ||    |   ||  |||||     | ||||     ||| ||    |
Db     130491 TGCCTGAATAATGGGTAAATAAACTATGGTTTGTCTGTAGATAAAATACTCATCAGCCGG 130550

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||  |   ||   | ||   |  | |  ||  | ||  |  |  |  |      |
Db     130551 TAAATATCATATTATATAAGAATATCAGATAACATAAGAAAATGTTCTCGGT-------- 130602

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
                |||   |||  |    | | |  |        |||||||| |||| ||||   ||| |
Db     130603 --ATATTGTTAGGTAGAAAGAAAAATCAAGTTACGGCCAGGCACAGTGGCTCACACCTAT 130660

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| ||||||||||||||  |||  ||||||||||  ||||||||||||  |||||
Db     130661 AATCCCAGCACTTTGGGAGGCTGAGGCAGGCGGATCACTGGAGGTCAGGAGTTCGAGACC 130720

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||| |||||||| |||||||||| |||||||||||||||| ||||||||||||||   |
Db     130721 AGCTTGGCCAACGTGGTGAAACCCCGTCTCTACTAAAAACACAAAAATTAGCCAGGCACG 130780

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||  ||| |||||||||||| |||||||||  | |||||  |||||||| |||||||||
Db     130781 GTGGTGTGCGCCTGTAGTCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCGCTTGAACC 130840

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCTGG-CAACAG 656
              | |||||||||||||  | | | |||||   ||| ||||||||||||||||||   ||||
Db     130841 TGGGAGGTGGAGGTTTTACTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGATGACAG 130900

Qy        657 AGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTAC 710
              ||  |||||||||  || | || |||||  |   |||||  || | |      || ||
Db     130901 AGTGAGACTCCATTAAAAAAAAAAAAAGTTACAAAATTATGCACAAATTATGCAC 130954
```

```
RESULT 34
AC027113
LOCUS       AC027113               159391 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 1 clone RP11-672I2 map 1, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC027113
VERSION     AC027113.3  GI:8081286
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 159391)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 1, clone RP11-672I2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159391)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
```

```
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 159391)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7656748.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8843
                Center clone name: 672_I_2
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149154 bases at least Q40
                Consensus quality: 154165 bases at least Q30
                Consensus quality: 156049 bases at least Q20
                Insert size: 195000; agarose-fp
                Insert size: 157491; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 5.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1373: contig of 1373 bp in length
            *     1374     1473: gap of 100 bp
            *     1474     3371: contig of 1898 bp in length
            *     3372     3471: gap of 100 bp
            *     3472     5991: contig of 2520 bp in length
            *     5992     6091: gap of 100 bp
            *     6092     8001: contig of 1910 bp in length
            *     8002     8101: gap of 100 bp
            *     8102    10452: contig of 2351 bp in length
            *    10453    10552: gap of 100 bp
            *    10553    13499: contig of 2947 bp in length
            *    13500    13599: gap of 100 bp
            *    13600    17300: contig of 3701 bp in length
            *    17301    17400: gap of 100 bp
            *    17401    23167: contig of 5767 bp in length
            *    23168    23267: gap of 100 bp
            *    23268    27394: contig of 4127 bp in length
            *    27395    27494: gap of 100 bp
            *    27495    31941: contig of 4447 bp in length
            *    31942    32041: gap of 100 bp
            *    32042    36867: contig of 4826 bp in length
            *    36868    36967: gap of 100 bp
            *    36968    43237: contig of 6270 bp in length
            *    43238    43337: gap of 100 bp
            *    43338    52795: contig of 9458 bp in length
            *    52796    52895: gap of 100 bp
            *    52896    61446: contig of 8551 bp in length
            *    61447    61546: gap of 100 bp
            *    61547    68822: contig of 7276 bp in length
            *    68823    68922: gap of 100 bp
            *    68923    79907: contig of 10985 bp in length
            *    79908    80007: gap of 100 bp
            *    80008    91003: contig of 10996 bp in length
            *    91004    91103: gap of 100 bp
            *    91104   109428: contig of 18325 bp in length
            *   109429   109528: gap of 100 bp
            *   109529   135277: contig of 25749 bp in length
            *   135278   135377: gap of 100 bp
            *   135378   159391: contig of 24014 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159391
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1"
                     /clone="RP11-672I2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1373
                     /note="assembly_fragment"
     gap             1374. .1473
                     /estimated_length=100
     misc_feature    1474. .3371
                     /note="assembly_fragment"
     gap             3372. .3471
                     /estimated_length=100
     misc_feature    3472. .5991
                     /note="assembly_fragment"
     gap             5992. .6091
                     /estimated_length=100
     misc_feature    6092. .8001
                     /note="assembly_fragment"
     gap             8002. .8101
                     /estimated_length=100
     misc_feature    8102. .10452
                     /note="assembly_fragment"
     gap             10453. .10552
                     /estimated_length=100
     misc_feature    10553. .13499
                     /note="assembly_fragment"
     gap             13500. .13599
                     /estimated_length=100
     misc_feature    13600. .17300
                     /note="assembly_fragment"
     gap             17301. .17400
                     /estimated_length=100
     misc_feature    17401. .23167
                     /note="assembly_fragment"
     gap             23168. .23267
                     /estimated_length=100
     misc_feature    23268. .27394
                     /note="assembly_fragment
```

```
                clone_end:T7
                vector_side:right"
    gap         27395. .27494
                /estimated_length=100
    misc_feature 27495. .31941
                /note="assembly_fragment"
    gap         31942. .32041
                /estimated_length=100
    misc_feature 32042. .36867
                /note="assembly_fragment"
    gap         36868. .36967
                /estimated_length=100
    misc_feature 36968. .43237
                /note="assembly_fragment"
    gap         43238. .43337
                /estimated_length=100
    misc_feature 43338. .52795
                /note="assembly_fragment"
    gap         52796. .52895
                /estimated_length=100
    misc_feature 52896. .61446

 Query Match            11.0%;  Score 219.2;  DB 14;  Length 159391;
 Best Local Similarity  54.8%;  Pred. No. 1.6e-35;
 Matches  390;  Conservative    0;  Mismatches  303;  Indels   19;  Gaps    3;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||  ||  ||||||||||||||||||||||||||| |||| ||||||||   || ||
Db      79738 TGATCATCCCACCTTGGCCTCCCAAAGTGCTGGGATTATAGGCATGAGCCACTGTGCTTG 79797

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |   || || ||   ||   | |   ||   |      | |   | |     |  ||
Db      79798 GCC---TGGCTGTAAACTTACTACATGGAAGTAGAAAAAAATAGAAGGCAGAGTTAGGAG 79854

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | |        || |  |||||  |  |||  |||  ||    |  |||| |||
Db      79855 ACAGGAAACCTGATTTTAATCCTGACTCTGATACTAGCTATGCAACCTTGAATNNNNNNN 79914

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
Db      79915 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 79974

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
                                                |  |  | | |     ||||
Db      79975 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGAACATTTCAAGGCAATAACAATACT 80034

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |   |   | | ||| | |  |||| | |   || ||  |  |  ||
Db      80035 ACACAAATTAAAAGCAAGTAAAAAATTATTATAAACAGGTGTTGATTCATCTAGGAA-- 80091

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
                  ||  |  | ||     | ||| | |||||||  |  || |||| |||||||
Db      80092 ----------AAGATTGTGAAAGAGGTAGAACTTGGCCAGGTGTGGTGGCTCATGCCTGT 80141

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||| |||||||||||||||||||||||||||  |||||
Db      80142 AATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACC 80201

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||| ||||||| |||||||| || ||||||||||||||   ||||||||||  ||  ||
Db      80202 AGCATGGCCAATATGGTGAAAGCCCGTCTCTACTAAAAATAAAAAAATTAGCTGGGCATG 80261

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              |||||| |||||||| |||||||||||| |||||||||| |||||||| |||||||||
Db      80262 GTGGCAGGCACCTGTAATCCCAACTACTCGGGAGGTTGAGGCAGGAGAATCGCTTGAACC 80321

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCTGGCAACAGA 657
              |||||  |||||||||||| | || ||    ||| || || |||||  |||||      |
Db      80322 CGGGAGGCGGAGGTTGCAGTGAGCCAAGATCATGCCATTGTACTCCAACCTGGGGGACAA 80381

Qy        658 GCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTA 709
              |  ||  |  |     ||| ||||  ||  |||      ||| ||   |||
Db      80382 GAGCGAAACTTTGTCTCCAAAAAAAATTTAAAAAAAAAAAAAAAGAAAGGTA 80433
```

```
RESULT 35
AC017038
LOCUS       AC017038              180707 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-313B2, WORKING DRAFT SEQUENCE,
            20 unordered pieces.
ACCESSION   AC017038
VERSION     AC017038.6  GI:8568131
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 180707)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180707)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:7230978.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0313B02
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; %
            Sequencing vector: plasmid; %
            Chemistry: Dye-primer ET; % of reads
            Chemistry: Dye-terminator Big Dye; % of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 168786 bases at least Q40
            Consensus quality: 172458 bases at least Q30
            Consensus quality: 174496 bases at least Q20
            Insert size: 200000; agarose-fp
            Insert size: 178807; sum-of-contigs
            Quality coverage: 3.48 in Q20 bases; agarose-fp
            Quality coverage: 3.86 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2076: contig of 2076 bp in length
            *     2077     2176: gap of unknown length
            *     2177     4769: contig of 2593 bp in length
            *     4770     4869: gap of unknown length
            *     4870     7558: contig of 2689 bp in length
            *     7559     7658: gap of unknown length
            *     7659    10945: contig of 3287 bp in length
            *    10946    11045: gap of unknown length
            *    11046    13945: contig of 2900 bp in length
            *    13946    14045: gap of unknown length
            *    14046    17724: contig of 3679 bp in length
            *    17725    17824: gap of unknown length
            *    17825    22156: contig of 4332 bp in length
            *    22157    22256: gap of unknown length
            *    22257    26042: contig of 3786 bp in length
            *    26043    26142: gap of unknown length
            *    26143    32662: contig of 6520 bp in length
            *    32663    32762: gap of unknown length
```

```
            *    32763    41257: contig of 8495 bp in length
            *    41258    41357: gap of unknown length
            *    41358    47568: contig of 6211 bp in length
            *    47569    47668: gap of unknown length
            *    47669    54391: contig of 6723 bp in length
            *    54392    54491: gap of unknown length
            *    54492    62210: contig of 7719 bp in length
            *    62211    62310: gap of unknown length
            *    62311    70009: contig of 7699 bp in length
            *    70010    70109: gap of unknown length
            *    70110    78968: contig of 8859 bp in length
            *    78969    79068: gap of unknown length
            *    79069    90881: contig of 11813 bp in length
            *    90882    90981: gap of unknown length
            *    90982   105415: contig of 14434 bp in length
            *   105416   105515: gap of unknown length
            *   105516   121842: contig of 16327 bp in length
            *   121843   121942: gap of unknown length
            *   121943   141936: contig of 19994 bp in length
            *   141937   142036: gap of unknown length
            *   142037   180707: contig of 38671 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180707
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-313B2"
     misc_feature    1. .2076
                     /note="assembly_name:Contig8"
     gap             2077. .2176
                     /estimated_length=unknown
     misc_feature    2177. .4769
                     /note="assembly_name:Contig9"
     gap             4770. .4869
                     /estimated_length=unknown
     misc_feature    4870. .7558
                     /note="assembly_name:Contig10"
     gap             7559. .7658
                     /estimated_length=unknown
     misc_feature    7659. .10945
                     /note="assembly_name:Contig11"
     gap             10946. .11045
                     /estimated_length=unknown
     misc_feature    11046. .13945
                     /note="assembly_name:Contig12"
     gap             13946. .14045
                     /estimated_length=unknown
     misc_feature    14046. .17724
                     /note="assembly_name:Contig13"
     gap             17725. .17824
                     /estimated_length=unknown
     misc_feature    17825. .22156
                     /note="assembly_name:Contig14"
     gap             22157. .22256
                     /estimated_length=unknown
     misc_feature    22257. .26042
                     /note="assembly_name:Contig15"
     gap             26043. .26142
                     /estimated_length=unknown
     misc_feature    26143. .32662
                     /note="assembly_name:Contig16"
     gap             32663. .32762
                     /estimated_length=unknown
     misc_feature    32763. .41257
                     /note="assembly_name:Contig17"
     gap             41258. .41357
                     /estimated_length=unknown
     misc_feature    41358. .47568
                     /note="assembly_name:Contig18"
     gap             47569. .47668
                     /estimated_length=unknown
     misc_feature    47669. .54391
                     /note="assembly_name:Contig19"
     gap             54392. .54491
                     /estimated_length=unknown
     misc_feature    54492. .62210
                     /note="assembly_name:Contig20"
     gap             62211. .62310
                     /estimated_length=unknown
     misc_feature    62311. .70009
                     /note="assembly_name:Contig21"
     gap             70010. .70109
                     /estimated_length=unknown
     misc_feature    70110. .78968
                     /note="assembly_name:Contig22"
     gap             78969. .79068
                     /estimated_length=unknown
     misc_feature    79069. .90881
                     /note="assembly_name:Contig23"
     gap             90882. .90981
                     /estimated_length=unknown
     misc_feature    90982. .105415
                     /note="assembly_name:Contig24
                     clone_end:SP6
                     vector_side:right"
     gap             105416. .105515
                     /estimated_length=unknown
     misc_feature    105516. .121842
                     /note="assembly_name:Contig25"
     gap             121843. .121942
                     /estimated_length=unknown
     misc_feature    121943. .141936
                     /note="assembly_name:Contig26"
     gap             141937. .142036
                     /estimated_length=unknown
     misc_feature    142037. .180707
                     /note="assembly_name:Contig27"
ORIGIN

  Query Match           11.0%;  Score 219.2;  DB 14;  Length 180707;
  Best Local Similarity  72.1%;  Pred. No. 1.6e-35;
  Matches  300;  Conservative    0;  Mismatches  113;  Indels    3;  Gaps    1;

Qy        351 TTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGC 410
              | |  |||  ||| ||  ||  |  ||| ||     | | |||  ||||||||||||||||
Db     139975 TCTGTAAATTTAAAACTATTCTAAGATTTAAATTTTCTTAAAGAGGCCAGGCGCAGTAGC 140034

Qy        411 TCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGA 470
              || ||||||||||||| ||||||  ||||||  |||||||  |||||||  |||||||||
Db     140035 TCATGCCTGTAATCCCAGCACTTTAGGAGGCTGAGGTGGGTGGATCACGGGAGGTCAGGA 140094

Qy        471 GTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTA 530
              |||  |||||||  |||||||||||||||||||  | |||||||||||||  |||||||||
Db     140095 GTTCAAGACCATCCTGGCCAACATGGTGAAATACAGTCTCTACTAAAAATACAAAAATTA 140154

Qy        531 GCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAAT 590
              ||  ||  |||||||| |  |||  || |||||  ||||||  |||  |  |||||  ||||||||
Db     140155 GCTGGGCATGGTGGCAGGTGCCTATAATCCCAGCTACTCGGGAAGCTGAGGCAGGAGAAT 140214

Qy        591 TGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCC 647
              ||||||||  |||||||||||||||||||| | ||||||   ||  ||||||||||||||
Db     140215 CACTTGAACCCAGGAGGTGGAGGTTGCAGTGAGCCGAGATGGTGCCACTGCACTCCAGCC 140274

Qy        648 TGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTG 707
              ||||  |||||||||||||||||     | ||  | ||    |  ||||       ||    ||
Db     140275 TGGCGACAGAGCAAGACTCCATCTCAAAAAAATAAAAATAAAAATAAAAATAAAAGTTTA 140334

Qy        708 TACCTATCTTCATTTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAG 763
              |       |   || |      ||    | |    | | || | |||  |    | ||||
Db     140335 TTAAAGAATAAGTTAATATGCTTATCCCTCAATAAATTCTATTTCATCTTCGTGAG 140390


RESULT 36
AC025682/c
LOCUS       AC025682              158135 bp    DNA     linear   PRI 01-FEB-2002
```

```
DEFINITION  Homo sapiens chromosome 17, clone RP11-580I16, complete sequence.
ACCESSION   AC025682
VERSION     AC025682.6  GI:18377318
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 158135)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-580I16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158135)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 158135)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 27, 2002 this sequence version replaced gi:18370031.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8312
                Center clone name: 580_I_16
            --------------.
FEATURES             Location/Qualifiers
     source          1. .158135
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-580I16"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(4. .215)
                     /rpt_family="AluJb"
     repeat_region   complement(237. .484)
                     /rpt_family="AluJb"
     repeat_region   complement(500. .801)
                     /rpt_family="AluSq"
     repeat_region   complement(2336. .2640)
                     /rpt_family="AluSx"
     repeat_region   3200. .3231
                     /rpt_family="MER5A"
     repeat_region   3232. .3501
                     /rpt_family="AluSg1"
     repeat_region   3502. .3588
                     /rpt_family="MER5A"
     repeat_region   5709. .5744
                     /rpt_family="(A)n"
     repeat_region   6191. .6237
                     /rpt_family="G-rich"
     unsure          6197. .6204
                     /note="<30 qual SNGL region"
     unsure          6206. .6211
                     /note="probably GGGGGG; possibly GGGGG"
     repeat_region   6597. .6652
                     /rpt_family="G-rich"
     unsure          complement(6636. .6640)
                     /note="<30 qual SNGL region"
     unsure          complement(6676. .6681)
                     /note="<30 qual SNGL region"
     repeat_region   6847. .6873
                     /rpt_family="(GGGA)n"
     repeat_region   complement(7516. .7665)
                     /rpt_family="MIR3"
     repeat_region   complement(7646. .7737)
                     /rpt_family="MIR"
     repeat_region   8051. .8345
                     /rpt_family="AluSc"
     repeat_region   8360. .8463
                     /rpt_family="FLAM_C"
     repeat_region   8504. .8807
                     /rpt_family="AluSx"
     repeat_region   9207. .9218
                     /rpt_family="AluS"
     repeat_region   9219. .9511
                     /rpt_family="AluSx"
     repeat_region   9512. .9574
                     /rpt_family="AluS"
     repeat_region   9575. .9627
                     /rpt_family="Alu"
     repeat_region   9640. .9917
                     /rpt_family="AluSx"
     repeat_region   complement(10167. .10285)
                     /rpt_family="AluSg/x"
     repeat_region   complement(10396. .10571)
                     /rpt_family="AluSq"
     repeat_region   10611. .10708
                     /rpt_family="FRAM"
     repeat_region   10865. .11044
                     /rpt_family="MIR"
     repeat_region   11052. .11348
                     /rpt_family="AluSx"
```

```
     repeat_region   11358. .11387
                     /rpt_family="AT_rich"
     repeat_region   complement(11592. .11957)
                     /rpt_family="L2"
     repeat_region   complement(11999. .12068)
                     /rpt_family="Alu"
     repeat_region   complement(12276. .12520)
                     /rpt_family="AluJb"
     repeat_region   complement(12734. .13039)
                     /rpt_family="AluJb"
     repeat_region   complement(13520. .13787)
                     /rpt_family="AluSx"
     repeat_region   complement(13792. .13986)
                     /rpt_family="MER20"
     repeat_region   complement(14103. .14403)
                     /rpt_family="AluSp"
     repeat_region   complement(14405. .14714)
                     /rpt_family="AluJb"
     repeat_region   complement(14715. .14847)
                     /rpt_family="MIR"
     repeat_region   14848. .15167
                     /rpt_family="AluSg"
     repeat_region   complement(15168. .15300)
                     /rpt_family="MIR"
     repeat_region   15415. .15724
                     /rpt_family="AluSp"
     repeat_region   complement(15756. .15884)
                     /rpt_family="L1ME4A"
     repeat_region   complement(15897. .16025)
                     /rpt_family="L1ME2"
     repeat_region   complement(16051. .16245)
                     /rpt_family="MER63A"
     repeat_region   complement(16251. .16675)
                     /rpt_family="L1ME1"
     repeat_region   complement(16676. .16938)
                     /rpt_family="AluJb"
     repeat_region   complement(16945. .17019)
                     /rpt_family="L1ME1"
     repeat_region   complement(17073. .17206)
                     /rpt_family="L1ME1"
     repeat_region   complement(17207. .17511)
                     /rpt_family="AluJb"
     repeat_region   complement(17512. .17783)
                     /rpt_family="L1ME1"
     repeat_region   complement(17784. .18053)
                     /rpt_family="AluSx"
     repeat_region   complement(18054. .18318)
                     /rpt_family="L1ME1"
     repeat_region   complement(18319. .18615)
                     /rpt_family="AluY"
     repeat_region   complement(18616. .18698)
                     /rpt_family="L1ME1"
     repeat_region   18717. .18888

  Query Match             10.9%;  Score 219;  DB 8;  Length 158135;
  Best Local Similarity   70.5%;  Pred. No. 1.8e-35;
  Matches  321;  Conservative    0;  Mismatches  130;  Indels    4;  Gaps    2;

Qy        272 TGATATGAAATAAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTATTTA 331
              || |   |||   | |     || | | | | ||     | | |  | |||| |    |
Db     145613 TGGTTAAAAAGTTGCACAGACACTATTTGAACCTAGACATTAGGTCTCTAGAGCTTCACA 145554

Qy        332 TTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAA 391
              |   | |   ||   | | |   | |||| |  |||||  ||  ||    | | | |
Db     145553 CTCTCAATCCCTACAGACACATCTCAGACTAGTCTACTTAGGTTCATATCAAATTCTAAT 145494

Qy        392 AT-GGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGG 450
              |   ||||||| |||||||| ||||||| ||||||||||||||||||  ||| |||
Db     145493 CTAACGCCAGGCACAGTAGCTCATGCCTGTTATCCCAACACTTTGGGAGGCTGAGGCGGG 145434

Qy        451 CGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTC 510
              |||||||||||||||||||| |||||||||||| |||||||||||||||||||| |||
Db     145433 AGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGTCAACATGGTGAAACCCTGCCTC 145374

Qy        511 TACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCA 570
              |||||||||  |||||||||||||||||||||||||  |||||||||||||| ||||||
Db     145373 TACTAAAAATACAAAAATTAGCCAGGTGTGGTGGCACACACCTGTAGTCCCAGCTACTCG 145314

Qy        571 GGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA- 629
              ||||| |||| ||||||||  ||||||||| |||||||||||||||||||| | ||||||
Db     145313 GGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCGAGAT 145254

Qy        630 --TGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTT 687
                || |||||||||||||||||  |||||| ||| |||       | || ||||
Db     145253 CGTGCCACTGCACTCCAGCCTGGGGACAGAGCGAGATTCCGCCTCAAAAAAAAAAAAAAA 145194

Qy        688 GAAATTGTGTAAATGAGTTGTACCTATCTTCATTT 722
              |||      |||  |     ||  ||||   |||
Db     145193 AAAAAGCAAAAAACAACACAAACTAATCTAACTTT 145159


RESULT 37
AC153308
LOCUS       AC153308               248748 bp    DNA     linear   HTG 23-JUL-2005
DEFINITION  Gorilla gorilla gorilla clone CH255-200A8, WORKING DRAFT SEQUENCE,
            5 ordered pieces.
ACCESSION   AC153308
VERSION     AC153308.3  GI:71067174
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Gorilla gorilla gorilla (lowland gorilla)
  ORGANISM  Gorilla gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 248748)
  AUTHORS   Antonellis,A., Ayele,K., Bass,D., Benjamin,B., Bera,J.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G.,
            Coleman,H., Engle,J., Franks,S., Fuksenko,T., Gestole,M.,
            Greene,A., Guan,X., Gupta,J., Gurson,N., Haghighi,P., Han,J.,
            Hansen,N., Ho,S.-L., Hu,P., Hunter,G., Hurle,B., Idol,J.R.,
            Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q., Legaspi,R.,
            Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Mojidi,H.A., Mullikin,J.C., Park,M.,
            Portnoy,M.E., Prasad,A., Puri,O., Rantz,K., Reddix-Dugue,N.,
            Sante,A., Schandler,K., Schueler,M.G., Sison,C., Stantripop,S.,
            Taye,A., Thomas,J.W., Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L.,
            Wetherby,K.D., Withers,T.R., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 248748)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-2004) NIH Intramural Sequencing Center, 5625
            Fishers Lane, Rockville, MD 20852, USA
REFERENCE   3  (bases 1 to 248748)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-2005) NIH Intramural Sequencing Center, 5625
            Fishers Lane, Rockville, MD 20852, USA
COMMENT     On Jul 23, 2005 this sequence version replaced gi:59676666.
            ---------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ---------- Project Information
                 Center project name: eqq
                 Center clone name: 200A08

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
```

```
          the sequence assembly is generally based on at least 8X average
          coverage in Q20 bases and has been reviewed to rule out
          gross misassemblies, the low-quality ends of sequence
          contigs have been trimmed away, and each base is associated
          with a Phrap-derived quality score.
          ---------- Summary Statistics
               Sequencing vector: plasmid; n/a; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.990319
               Consensus quality: 246798 bases at least Q40
               Consensus quality: 247845 bases at least Q30
               Consensus quality: 248232 bases at least Q20
               Insert size: 212000; agarose-fp
               Insert size: 248348; sum-of-contigs
               Quality coverage: 8.53x in Q20 bases; agarose-fp
               Quality coverage: 7.28x in Q20 bases; sum-of-contigs
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 5 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1    55010: contig of 55010 bp in length
          *    55011    55110: gap of unknown length
          *    55111    75860: contig of 20750 bp in length
          *    75861    75960: gap of unknown length
          *    75961    94951: contig of 18991 bp in length
          *    94952    95051: gap of unknown length
          *    95052   153305: contig of 58254 bp in length
          *   153306   153405: gap of unknown length
          *   153406   248748: contig of 95343 bp in length.
FEATURES             Location/Qualifiers
     source          1. .248748
                     /organism="Gorilla gorilla gorilla"
                     /mol_type="genomic DNA"
                     /sub_species="gorilla"
                     /db_xref="taxon:9595"
                     /clone="CH255-200A8"
                     /clone_lib="CH255"
                     /note="BAC resource: http://bacpac.chori.org/"
     misc_feature    1. .110099
                     /note="clone overlaps with GenBank Accession Number
                     AC153309 clone CH255-214B7 (center project name eqr)"
     misc_feature    1. .55010
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     gap             55011. .55110
                     /estimated_length=unknown
     misc_feature    55111. .75860
                     /note="assembly_fragment"
     gap             75861. .75960
                     /estimated_length=unknown
     misc_feature    75961. .94951
                     /note="assembly_fragment"
     gap             94952. .95051
                     /estimated_length=unknown
     misc_feature    95052. .153305
                     /note="assembly_fragment"
     gap             153306. .153405
                     /estimated_length=unknown
     misc_feature    153406. .248748
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             10.9%;  Score 219;  DB 14;  Length 248748;
  Best Local Similarity   86.9%;  Pred. No. 1.6e-35;

  Matches  253;  Conservative    0;  Mismatches   35;  Indels    3;  Gaps    1;

Qy        395 GGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGA 454
              ||||||||||||| |||| |||||||||||||| |||||||||||||||||||||| ||||
Db     125525 GGCCAGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGACGGA 125584

Qy        455 TCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACT 514
              | || ||||||| ||||||| |||||| ||||||||||||||||||||||||||||||||
Db     125585 TTACTTGAGGTCGGGAGTTTGAGACCATCCTGGCCAACATGGTGAAACCCTGTCTCTACT 125644

Qy        515 AAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAG 574
              |||||  |||||||||||||||  |||| ||| || |||||| ||||||||||||||||
Db     125645 AAAAATACAAAAATTAGCCAGGCATGGTTGCAGGCGCCTGTAATCCCAACTACTCAGGAG 125704

Qy        575 GTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---G 631
              ||||||| |||||||||||||||||| ||||| |||||||||||| | | |||||   |
Db     125705 GTTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTGAGATCGCG 125764

Qy        632 TCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              ||||| | |||||||||| ||||||| |||||||||    | || ||||
Db     125765 CCACTGTATTCCAGCCTGGTGACAGAGCGAGACTCCATCTCAAAAAGAAAA 125815


RESULT 38
AC004079
LOCUS       AC004079               102717 bp    DNA     linear   PRI 27-JAN-2004
DEFINITION  Homo sapiens PAC clone RP1-167F23 from 7, complete sequence.
ACCESSION   AC004079
VERSION     AC004079.1  GI:2822174
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 102717)
  AUTHORS   Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
            Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
            Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
            Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
            Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
            Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L.,
            Kalicki,J., Ozersky,P., Bielicki,L., Scott,K., Holmes,A.,
            Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
            Dauphin-Kohlberg,S., Kozlowicz-Reilly,A., Leonard,S., Rohlfing,T.,
            Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
            Strowmatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
            Woessner,J.P., Wendl,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,
            Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohldmann,P.E.,
            Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A.,
            Mardis,E.R., Clifton,S.W., Chissoe,S.L., Marra,M.A., Raymond,C.,
            Haugen,E., Gillett,W., Zhou,Y., James,R., Phelps,K., Iadanoto,S.,
            Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J.,
            Furey,T.S., Baertsch,R.A., Brent,M.R., Keibler,E., Flicek,P.,
            Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrents,D.,
            Chinwalla,A.T., Gish,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V.,
            Eichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K.
  TITLE     The DNA sequence of human chromosome 7
  JOURNAL   Nature 424 (6945), 157-164 (2003)
   PUBMED   12853948
REFERENCE   2  (bases 1 to 102717)
  AUTHORS   Duckels,G., Hawkins,M., Hinds,K. and Jones,K.
  TITLE     The sequence of Homo sapiens PAC clone RP1-167F23
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 102717)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 102717)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
```

```
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 102717)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 102717)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 102717)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2004) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_DJ0167F23
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This clone was provided for sequencing by Dr. Stephen Scherer,
            Department of Genetics, The Hospital for Sick Children, Toronto,
            Ontario, Canada, with support from the Canadian Genome Analysis and
            Technology Program; and Dr. John D. McPherson, Department of
            Genetics, Washington University, St. Louis MO. For additional
            information about the map position of this sequence, see
            http://www.genet.sickkids.on.ca/chromosome7 and
            http://genome.wustl.edu

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1, prepared by
            Pieter de Jong and coworkers at http://www.chori.org using the
            method described by Ioannou et al., Nature Genetics 6:84-9 (1994).
            The library is from one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2

            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of RP1-167F23
            the actual end is at base position 102717 of RP1-167F23. The
            orientation of this clone is unknown.
FEATURES                Location/Qualifiers
     source             1. .102717
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="7"
                        /map="7"
                        /clone="RP1-167F23"
                        /clone_lib="RPCI-1"
     repeat_region      1. .1098
                        /rpt_family="L1"
     repeat_region      1099. .1412
                        /rpt_family="Alu"
     repeat_region      1413. .2186
                        /rpt_family="L1"
     repeat_region      2187. .2495
                        /rpt_family="Alu"
     repeat_region      2496. .2910
                        /rpt_family="L1"
     repeat_region      3131. .3332
                        /rpt_family="Alu"
     repeat_region      3333. .3696
                        /rpt_family="MaLR"
     repeat_region      3697. .3799
                        /rpt_family="Alu"
     repeat_region      4056. .4081
                        /rpt_family="(TTTC)n"
     repeat_region      4145. .4277
                        /rpt_family="L2"
     repeat_region      4926. .5064
                        /rpt_family="MER1_type"
     repeat_region      5081. .5332
                        /rpt_family="L1"
     repeat_region      6427. .6546
                        /rpt_family="Alu"
     repeat_region      6572. .6868
                        /rpt_family="Alu"
     repeat_region      7942. .8236
                        /rpt_family="Alu"
     repeat_region      8544. .8579
                        /rpt_family="AT_rich"
     repeat_region      8991. .9300
                        /rpt_family="Alu"
     repeat_region      10140. .10406
                        /rpt_family="L1"
     repeat_region      10407. .10702
                        /rpt_family="Alu"
     repeat_region      10703. .10790
                        /rpt_family="L1"
     repeat_region      11252. .11503
                        /rpt_family="MIR"
     repeat_region      11814. .11854
                        /rpt_family="MIR"
     repeat_region      12097. .12443
                        /rpt_family="Alu"
     repeat_region      12503. .12554
                        /rpt_family="MIR"
     repeat_region      12555. .12687
                        /rpt_family="Alu"
     repeat_region      12705. .12783
                        /rpt_family="MIR"
     repeat_region      13560. .13859
                        /rpt_family="Alu"
     repeat_region      13863. .13976
                        /rpt_family="MIR"
     repeat_region      14007. .14163
                        /rpt_family="MIR"
     repeat_region      14385. .14604
                        /rpt_family="MIR"
     repeat_region      15123. .15178
                        /rpt_family="L2"
     repeat_region      15754. .16057
                        /rpt_family="Alu"
     repeat_region      16104. .16126
                        /rpt_family="AT_rich"
     repeat_region      16134. .16443
                        /rpt_family="Alu"
     repeat_region      16868. .16975
                        /rpt_family="MIR"
     repeat_region      16995. .17352
                        /rpt_family="MaLR"
     repeat_region      17356. .17434
                        /rpt_family="L1"
     repeat_region      17435. .17736
```

```
                     /rpt_family="Alu"
     repeat_region   17737. .18018
                     /rpt_family="Alu"
     repeat_region   18019. .18495
                     /rpt_family="L1"
     repeat_region   18498. .18788

  Query Match             10.9%;  Score 218.8;  DB 8;  Length 102717;
  Best Local Similarity   61.4%;  Pred. No. 2.1e-35;
  Matches  421;  Conservative    0;  Mismatches  257;  Indels    8;  Gaps    4;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||| || |||| |||||||| |||||||||||||||| |||||||||||||||| |
Db      52054 TGATCCGCCCGCCTCGGCCTCCCTAAGTGCTGGGATTACAAGCGTGAGCCACCACGCCAG 52113

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  | |  |  || ||||  | |  ||   | |  |     |  |         |
Db      52114 GCC--AGGGTTGAATAATTTTTAAATGCATTTTACACTTGGAGAAGATTGTACCAAAATC 52171

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              |  | ||      |||      |    || | |   |   |  |  |    || |
Db      52172 ATCCATTTGCATTTTCAGTGAATGACATATAAACATTTTAGTCATTATATATAAGCTGAA 52231

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
                | |   |    | ||| |  || | |   ||  |   |   |   |||    ||
Db      52232 AGTATAAAAAGATTGTTTATATATTTTTTTAGTATATCTACTTTTCATTTGTGAGAACCC 52291

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              || |   |  |  |     |  | ||              ||||    | || ||| | ||
Db      52292 TAGAGGAAAAAGTGTTACAAAAAAGAAAACACCCCCCAAAATCTCTACTTTAACGAATAA 52351

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
               ||||  |   |  ||  |  |    ||||  |||   |   | |||     ||     |
Db      52352 AAGCTAATGAAGTAGGCTGCATTCCTATTACTTAGTTCCTCTAAATGTCTATTCCCTTAT 52411

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              |||||    |  ||||||||     |      | |||| |||   || |||    |||||
Db      52412 TAATAACTCTGCTTTATTAAAATAACT--GTCTCGGCCGGGCATGGTGGCTGACACCTGT 52469

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| ||||||||||||  ||||||| |||||||||||||||| |||||| |||| |
Db      52470 AATCCCAGCACTTTGGGAGGCTGAGGTGGGTGGATCACCTGAGGTCGGGAGTTCAAGAGC 52529

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||| |||||||||||||||| | |||||||||| |||||||||| ||  ||
Db      52530 AGCCTGGCTAACATGGTGAAACCCTGCCACTACTAAAAACACAAAAATTAGCTGGGCATG 52589

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||   |||||||| | ||| ||||| |||| ||||| ||||||||||||||||||
Db      52590 GTGGTGGGCACCTGTAATTCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACC 52649

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACAG 656
              |||||||||||||||||| | || ||||   | ||||||||||||||| |||||||
Db      52650 CAGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCGCCACTGCACTCCAGCCTAGGCAACAG 52709

Qy        657 AGCAAGACTCCATAAAGACAACAAAA 682
              | | ||||||| |  | | || ||||
Db      52710 ACCGAGACTCCGTCTAAAAAAAAAAA 52735


RESULT 39
AC011380
LOCUS       AC011380               133955 bp    DNA     linear   PRI 01-NOV-2002
DEFINITION  Homo sapiens chromosome 5 clone CTB-133C10, complete sequence.
ACCESSION   AC011380
VERSION     AC011380.6  GI:24462307
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 133955)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133955)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 133955)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 1, 2002 this sequence version replaced gi:8576057.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .133955
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-133C10"
ORIGIN

  Query Match             10.9%;  Score 218.8;  DB 8;  Length 133955;
  Best Local Similarity   77.4%;  Pred. No. 2e-35;
  Matches  291;  Conservative    0;  Mismatches   82;  Indels    3;  Gaps    2;

Qy        372 AATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACAC 431
              |||| ||||| ||   | |    |||||||| |||| |||| ||||||||||||| |||
Db      26185 AATTCTTAAAAAGTGCTTATTATGGCCAGGCACAGTGGCTCATGCCTGTAATCCCAGCAC 26244

Qy        432 TTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAA 491
              |||||||||||||||| |||| ||||||||||||||||||| ||||||||||||||||
Db      26245 TTTGGGAGGCCAAGGCGGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAA 26304

Qy        492 CATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCAC 551
              |||||||||||||| ||||||||||||||  |||||| |||||||| |||||| | ||
Db      26305 CATGGTGAAACCCCGTCTCTACTAAAAATACAAAAACTAGCCAGGCGTGGTGGTACGCGT 26364

Qy        552 CTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGA 611
              || |||||||| |||||||||||| ||||| |||||||| |||||||||| |||||  ||
Db      26365 CTATAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATGGCTTGAACCTGGGAGGCAGA 26424

Qy        612 GGTTGCAGTAACCCGAGATGTCAC--TGCACTCCAGCCTGGCAACAGAGCAAGACTCCAT 669
              |||||| || | || || |  |||  ||||||| |||||||| ||||||  |||||||||
Db      26425 GGTTGCTGTGAGCCAAGGTCACACCATGCACTCTAGCCTGGCGACAGAGTGAGACTCCAT 26484

Qy        670 -AAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGAAA 728
               | | || ||||    |     |||||||      |       ||||  | | ||  | | |
Db      26485 CTCAAAAAAAAAAAAAAGTGCTTATTGTGTGTCACATGGTCTCCTAAGTGCTTTATATACA 26544

Qy        729 TTCATCTTTGTTCATT 744
              || |  |||  || ||
Db      26545 TTAACATTTAATCCTT 26560


RESULT 40
AC079931
LOCUS       AC079931               178459 bp    DNA     linear   HTG 16-MAR-2001
DEFINITION  Homo sapiens chromosome 7 clone CTD-2594L23, WORKING DRAFT
            SEQUENCE, 43 unordered pieces.
ACCESSION   AC079931
VERSION     AC079931.3  GI:13357557
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 178459)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178459)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 16, 2001 this sequence version replaced gi:11991333.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_TD2594L23
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 95%
            Sequencing vector: plasmid; 5%
            Chemistry: Dye-primer ET; 95% of reads
            Chemistry: Dye-terminator Big Dye; 5% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 154148 bases at least Q40
            Consensus quality: 162306 bases at least Q30
            Consensus quality: 166493 bases at least Q20
            Insert size: 191000; agarose-fp
            Insert size: 174259; sum-of-contigs
            Quality coverage: 3.14 in Q20 bases; agarose-fp
            Quality coverage: 3.53 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 43 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1135: contig of 1135 bp in length
            *     1136     1235: gap of unknown length
            *     1236     2532: contig of 1297 bp in length
            *     2533     2632: gap of unknown length
            *     2633     3791: contig of 1159 bp in length
            *     3792     3891: gap of unknown length
            *     3892     5003: contig of 1112 bp in length
            *     5004     5103: gap of unknown length
            *     5104     6609: contig of 1506 bp in length
            *     6610     6709: gap of unknown length
            *     6710     8108: contig of 1399 bp in length
            *     8109     8208: gap of unknown length
            *     8209     9889: contig of 1681 bp in length
            *     9890     9989: gap of unknown length
            *     9990    12104: contig of 2115 bp in length
            *    12105    12204: gap of unknown length
            *    12205    14028: contig of 1824 bp in length
            *    14029    14128: gap of unknown length
            *    14129    15404: contig of 1276 bp in length
            *    15405    15504: gap of unknown length
            *    15505    17820: contig of 2316 bp in length
            *    17821    17920: gap of unknown length
            *    17921    20527: contig of 2607 bp in length
            *    20528    20627: gap of unknown length
            *    20628    23434: contig of 2807 bp in length
            *    23435    23534: gap of unknown length
            *    23535    26463: contig of 2929 bp in length
            *    26464    26563: gap of unknown length
            *    26564    28733: contig of 2170 bp in length
            *    28734    28833: gap of unknown length
            *    28834    30672: contig of 1839 bp in length
            *    30673    30772: gap of unknown length
            *    30773    33394: contig of 2622 bp in length
            *    33395    33494: gap of unknown length
            *    33495    35974: contig of 2480 bp in length
            *    35975    36074: gap of unknown length
            *    36075    38223: contig of 2149 bp in length
            *    38224    38323: gap of unknown length
            *    38324    40744: contig of 2421 bp in length
            *    40745    40844: gap of unknown length
            *    40845    44797: contig of 3953 bp in length
            *    44798    44897: gap of unknown length
            *    44898    48446: contig of 3549 bp in length
            *    48447    48546: gap of unknown length
            *    48547    52596: contig of 4050 bp in length
            *    52597    52696: gap of unknown length
            *    52697    57018: contig of 4322 bp in length
            *    57019    57118: gap of unknown length
            *    57119    60603: contig of 3485 bp in length
            *    60604    60703: gap of unknown length
            *    60704    62021: contig of 1318 bp in length
            *    62022    62121: gap of unknown length
            *    62122    67532: contig of 5411 bp in length
            *    67533    67632: gap of unknown length
            *    67633    72176: contig of 4544 bp in length
            *    72177    72276: gap of unknown length
            *    72277    76066: contig of 3790 bp in length
            *    76067    76166: gap of unknown length
            *    76167    80719: contig of 4553 bp in length
            *    80720    80819: gap of unknown length
            *    80820    84399: contig of 3580 bp in length
            *    84400    84499: gap of unknown length
            *    84500    88672: contig of 4173 bp in length
            *    88673    88772: gap of unknown length
            *    88773    95190: contig of 6418 bp in length
            *    95191    95290: gap of unknown length
            *    95291   100594: contig of 5304 bp in length
            *   100595   100694: gap of unknown length
            *   100695   106783: contig of 6089 bp in length
            *   106784   106883: gap of unknown length
            *   106884   113874: contig of 6991 bp in length
            *   113875   113974: gap of unknown length
            *   113975   118494: contig of 4520 bp in length
            *   118495   118594: gap of unknown length
            *   118595   126676: contig of 8082 bp in length
            *   126677   126776: gap of unknown length
            *   126777   134408: contig of 7632 bp in length
            *   134409   134508: gap of unknown length
            *   134509   145575: contig of 11067 bp in length
            *   145576   145675: gap of unknown length
            *   145676   153565: contig of 7890 bp in length
            *   153566   153665: gap of unknown length
            *   153666   163958: contig of 10293 bp in length
            *   163959   164058: gap of unknown length
            *   164059   178459: contig of 14401 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178459
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="CTD-2594L23"
     misc_feature    1. .1135
                     /note="assembly_name:Contig32"
     gap             1136. .1235
                     /estimated_length=unknown
     misc_feature    1236. .2532
                     /note="assembly_name:Contig33"
     gap             2533. .2632
                     /estimated_length=unknown
     misc_feature    2633. .3791
                     /note="assembly_name:Contig34"
     gap             3792. .3891
                     /estimated_length=unknown
```

```
misc_feature    3892. .5003
                /note="assembly_name:Contig37"
gap             5004. .5103
                /estimated_length=unknown
misc_feature    5104. .6609
                /note="assembly_name:Contig38"
gap             6610. .6709
                /estimated_length=unknown
misc_feature    6710. .8108
                /note="assembly_name:Contig39"
gap             8109. .8208
                /estimated_length=unknown
misc_feature    8209. .9889
                /note="assembly_name:Contig43"
gap             9890. .9989
                /estimated_length=unknown
misc_feature    9990. .12104
                /note="assembly_name:Contig46"
gap             12105. .12204
                /estimated_length=unknown
misc_feature    12205. .14028
                /note="assembly_name:Contig47"
gap             14029. .14128
                /estimated_length=unknown
misc_feature    14129. .15404
                /note="assembly_name:Contig49"
gap             15405. .15504
                /estimated_length=unknown
misc_feature    15505. .17820
                /note="assembly_name:Contig50"
gap             17821. .17920
                /estimated_length=unknown
misc_feature    17921. .20527
                /note="assembly_name:Contig51"
gap             20528. .20627
                /estimated_length=unknown
misc_feature    20628. .23434
                /note="assembly_name:Contig52"
gap             23435. .23534
                /estimated_length=unknown
misc_feature    23535. .26463
                /note="assembly_name:Contig53"
gap             26464. .26563
                /estimated_length=unknown
misc_feature    26564. .28733
```

Query Match 10.9%; Score 218.8; DB 14; Length 178459;
Best Local Similarity 61.4%; Pred. No. 1.9e-35;
Matches 421; Conservative 0; Mismatches 257; Indels 8; Gaps 4;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||| || |||| |||||||| |||||||||||||||| |||||||||||||||| |
Db      68469 TGATCCGCCCGCCTCGGCCTCCCTAAGTGCTGGGATTACAAGCGTGAGCCACCACGCCAG 68528

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  | |  |  || ||||  | |  ||   |  | | |    | |     |    |
Db      68529 GCC--AGGGTTGAATAATTTTTAAATGCATTTTACACTTGGAGAAGATTGTACCAAAATC 68586

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              |  | ||      |||      |     || |  |   |   |  |    || |
Db      68587 ATCCATTTGCATTTTCAGTGAATGACATATAAACATTTTAGTCATTATATATAAGCTGAA 68646

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
                | |   |    | ||| |  || | |   ||  |   |   |   |||    ||
Db      68647 AGTATAAAAAGATTGTTTATATATTTTTTTAGTATATCTACTTTTCATTTGTGAGAACCC 68706

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              || |   |  |    |   | ||             ||||    | || ||| | ||
Db      68707 TAGAGGAAAAAGTGTTACAAAAAAGAAAACACCCCCCAAAATCTCTACTTTAACGAATAA 68766

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
               ||||  |   | ||  |  |   ||||  |||   |   | |||     ||    |
Db      68767 AAGCTAATGAAGTAGGCTGCATTCCTATTATTTAGTTCCTCTAAATGTCTATTCCCTTAT 68826

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              |||||    |   |||||||     |     | |||| |||   || |||    |||||
Db      68827 TAATAACTCTGCTTTATTAAAATAACT--GTCTCGGCCGGGCATGGTGGCTGACACCTGT 68884

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||  ||||||| |||||||||||||||| |||||| |||| |
Db      68885 AATCCCAGCACTTTGGGAGGCTGAGGTGGGTGGATCACCTGAGGTCGGGAGTTCAAGAGC 68944

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||| |||||||||||||||| | ||||||||||| ||||||||||   ||  ||
Db      68945 AGCCTGGCTAACATGGTGAAACCCTGCCACTACTAAAAACACAAAAATTAGCTGGGCATG 69004

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||   ||||||||| | ||| |||||| ||||| ||||| ||||||||||||||||||
Db      69005 GTGGTGGGCACCTGTAATTCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACC 69064

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACAG 656
              |||||||||||||||||| | || ||||   | ||||||||||||||| ||||||||
Db      69065 CAGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCGCCACTGCACTCCAGCCTGGGCAACAG 69124

Qy        657 AGCAAGACTCCATAAAGACAACAAAA 682
              | | ||||||| |  | | || ||||
Db      69125 ACCGAGACTCCGTCTAAAAAAAAAAA 69150
```

RESULT 41
AL136365/c

```
LOCUS       AL136365              179006 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-143M15 on chromosome 9p23-24.3
            Contains the DMRT1 gene for doublesex and mab-3 related
            transcription factor 1, the DMRT3 gene for doublesex and mab-3
            related transcription factor 3 and four CpG islands, complete
            sequence.
ACCESSION   AL136365
VERSION     AL136365.9  GI:8246868
KEYWORDS    HTG; CpG island; DMRT1; DMRT3; doublesex; mab-3.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 179006)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 4, 2000 this sequence version replaced gi:7801526.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            RP11-143M15 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
```

```
          one subclone; and the assembly was confirmed by restriction digest,
          except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .179006
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="p23-24.3"
                     /clone="RP11-143M15"
                     /clone_lib="RPCI-11.1"
     misc_feature    1
                     /note="Clone_left_end: RP11-143M15"
     gene            19544. .146944
                     /gene="DMRT1"
                     /locus_tag="RP11-143M15.1-001"
     mRNA            join(19544. .20046,24814. .24997,71766. .72049,94617. .94761,
                     145839. .146944)
                     /gene="DMRT1"
                     /locus_tag="RP11-143M15.1-001"
                     /product="doublesex and mab-3 related transcription factor
                     1"
                     /note="match: ESTs: AA412330.1 AA429597.1 AA432207.1
                     AI200978.1 AI203184.1 AI217960.1 AI698612.1 AI809359.1
                     AL043067 AW457053.1
                     match: cDNAs: AF123456 AF130728.1 AF192560.1 AF192561.1
                     AF202778.1 AF211349.1 AF216651.1 AJ276801.1 AL133300.1
                     AL162131.1"
     CDS             join(19693. .20046,24814. .24997,71766. .72049,94617. .94761,
                     145839. .145993)
                     /gene="DMRT1"
                     /locus_tag="RP11-143M15.1-001"
                     /standard_name="OTTHUMP00000020961"
                     /note="match: proteins: Q925A6 Q9QZ59 Q9TT01 Q9Y5R6"
                     /codon_start=1
                     /product="doublesex and mab-3 related transcription factor
                     1"
                     /protein_id="CAB99335.1"
                     /db_xref="GI:9453730"
                     /db_xref="Genew:2934"
                     /db_xref="GOA:Q9Y5R6"
                     /db_xref="HSSP:1LPV"
                     /db_xref="UniProt/Swiss-Prot:Q9Y5R6"
                     /translation="MPNDEAFSKPSTPSEAPHAPGVPPQGRAGGFGKASGALVGAASG
                     SSAGGSSRGGGSGSGASDLGAGSKKSPRLPKCARCRNHGYASPLKGHKRFCPWRDCQC
                     KKCNLIAERQRVMAAQVALRRQQAQEEELGISHPIPLPSAAELLVKRENNGSNPCLMT
                     ECSGTSQPPPASVPTTAASEGRMVIQDIPAVTSRGHVENTPDLVSDSTYYSSFYQPSL
                     FPYYNNLYNCPQYSMALAADSASGEVGNPLGGSPVKNSLRGLPGPYVPGQTGNQWQMK
                     NMENRHAMSSQYRMHSYYPPPSYLGQSVPQFFTFEDAPSYPEARASVFSPPSSQDSGL
                     VSLSSSSPISNKSTKAVLECEPASEPSSFTVTPVIEEDE"
     misc_feature    30905
                     /gene="DMRT1"
                     /locus_tag="RP11-143M15.1-001"
                     /note="Clone_right_end: RP11-130C19"
     polyA_signal    146920. .146925
                     /gene="DMRT1"
                     /locus_tag="RP11-143M15.1-001"
     polyA_site      146944
                     /gene="DMRT1"
                     /locus_tag="RP11-143M15.1-001"
     gene            join(154818. .155309,167895. .169585)
                     /gene="DMRT3"
                     /locus_tag="RP11-143M15.2-001"
     mRNA            join(154818. .155309,167895. .169585)
                     /gene="DMRT3"
                     /locus_tag="RP11-143M15.2-001"
                     /product="doublesex and mab-3 related transcription factor
                     3"
                     /note="match: ESTs: AA598944.1 AA627687.1 AI002773.1
                     AI052121.1 AI732189.1 AI733646.1 AV298122
                     match: cDNAs: AF193873.1 AJ301581.1"
     CDS             join(154856. .155309,167895. .168859)
                     /gene="DMRT3"
                     /locus_tag="RP11-143M15.2-001"
                     /standard_name="OTTHUMP00000020962"
                     /codon_start=1
                     /product="doublesex and mab-3 related transcription factor
                     3"
                     /protein_id="CAB99336.1"
                     /db_xref="GI:9453731"
                     /db_xref="Genew:13909"
                     /db_xref="GOA:Q9NQL9"
                     /db_xref="HSSP:1LPV"
                     /db_xref="InterPro:IPR001275"
                     /db_xref="InterPro:IPR005173"
                     /db_xref="UniProt/TrEMBL:Q9NQL9"
                     /translation="MNGYGSPYLYMGGPVSQPPRAPLQRTPKCARCRNHGVLSWLKGH
                     KRYCRFKDCTCEKCILIIERQRVMAAQVALRRQQANESLESLIPDSLRALPGPPPPGD
                     AVAAPQPPPASQPSQPQPPRPAAELAAAAALRWTAEPQPGALQAQLAKPDLTEERLGD
                     GKSADNTEVFSDKDTDQRSSPDVAKSKGCFTPESPEIVSVEEGGYAVQKNGGNPESRP
                     DSPKCHAEQNHLLIEGPSGTVSLPFSLKANRPPLEVLKKIFPNQKPTVLELILKGCGG
                     DLVSAVEVLLSSRSSVTGAERTSAEPESLALPSNGHIFEHTLSSYPISSSKWSVGSAF
                     RVPDTLRFSADSSNVVPSPLAGPLQPPFPQPPRYPLMLRNTLARSQSSPFLPNDVTLW
                     NTMTLQQQYQLRSQYVSPFPSNSTSVFRSSPVLPARATEDPRISIPDDGCPFVSKQSI
                     YTEDDYDERSDSSDSRTLNTSS"
     gene            167691. .>168187
                     /gene="DMRT3"
                     /locus_tag="RP11-143M15.2-002"
     mRNA            167691. .>168187
                     /gene="DMRT3"
                     /locus_tag="RP11-143M15.2-002"
                     /product="doublesex and mab-3 related transcription factor
                     3"
                     /note="match: ESTs: AW105228.1 AW592669.1
                     match: cDNAs: AF217741.2"
     CDS             167852. .>168187
                     /gene="DMRT3"
                     /locus_tag="RP11-143M15.2-002"
                     /standard_name="OTTHUMP00000020963"
                     /note="match: proteins: Q9NRJ9"
                     /codon_start=1
                     /product="doublesex and mab-3 related transcription factor
                     3"
                     /protein_id="CAH70546.1"
                     /db_xref="GI:55661454"
                     /db_xref="UniProt/TrEMBL:Q5W0Z5"
                     /translation="MHTRKRLTQLFCFPDLTEERLGDGKSADNTEVFSDKDTDQRSSP
                     DVAKSKGCFTPESPEIVSVEEGGYAVQKNGGNPESRPDSPKCHAEQNHLLIEGPSGTV
                     SLPFSLKANR"
     polyA_signal    169568. .169573
                     /gene="DMRT3"
                     /locus_tag="RP11-143M15.2-001"
     polyA_site      169586
     misc_feature    179006
                     /note="Clone_right_end: RP11-143M15"
ORIGIN

  Query Match             10.9%;  Score 218.8;  DB 8;  Length 179006;
  Best Local Similarity   60.7%;  Pred. No. 1.9e-35;
  Matches  465;  Conservative    0;  Mismatches  287;  Indels   14;  Gaps    6;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACG-CCT 59
              || ||  || |||||||||||||||| |||||||||| |||| |||||||||| || ||
Db     175382 TGGTCTGCCTGCCTTGGCCTCCCAAAATGCTGGGATTGCAGGTGTGAGCCACCGCGCCCA 175323

Qy         60 GGTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |||    ||| ||| |  || |      |||  |  ||    |      |  | ||| |
Db     175322 GCTAGAAGTGTTTTTTTAATAAAATAATTTGCATTTGGATTGGTTTCAGATAATGTTTCA 175263

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAA-CGGCCTTAAATGACTT 178
               |    |   | |       |  |||         | ||| || | |  ||  ||    |
Db     175262 AAATAATACCTAGCAGGTGCTTTGAATTACTAATTTGACTAAATCAGTGTTGGATTCATA 175203

Qy        179 AGTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGG 238
              ||  | ||||    |   |       |      | || ||       | ||||
Db     175202 AGATTTTCTCTAAGGGTATAAGCAGTGGTATACAAAGCATTATTCCAATTCTCAAAACTC 175143
```

```
Qy        239 AATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGT 298
              | ||      | |   || |   ||   |    ||   | | | | | | | | |
Db     175142 TGGATACAGTAAGATATCTTTAATACTAAGCTGCATTACACATAATGTTTCTTGAAGATT 175083

Qy        299 AGTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAA 358
                  | |||| |  || | |  | |||||  |   |  | | |   |
Db     175082 CAGGATCTGCAGTGAATCTCAAATATTACTGTATTAAGAGATATGCAGACAGGCTGGCTC 175023

Qy        359 ACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCT 418
              ||  |   |  |  ||  || |      |||| ||| || ||||| |||| | |||
Db     175022 ACAGAGCTATCTTGAAGATAGAAAACTGCATAAATCGGCTGGGTGCAGTGGCTCATACCT 174963

Qy        419 GTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGA 478
              |||||||||||||||| |||||||| ||| ||||||||||   |||||||||| | ||||
Db     174962 GTAATCCCAACACTTCGGGAGGCCGAGGCGGGCGGATCAC--AAGGTCAGGAGGTCAAGA 174905

Qy        479 CCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTG 538
              ||| |||||||||||| ||||||||| |||||||||||||  ||||||||||| ||||
Db     174904 CCATCCTGGCCAACATCGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCGGGTG 174845

Qy        539 TGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAA 598
              ||||||    |||||| ||||||| |||||||||||| ||||| ||||||||||||||
Db     174844 TGGTGGTGCACACCTGCAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAA 174785

Qy        599 CCTAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCAC---TGCACTCCAGCCTGGCAACA 655
              ||  ||||| | |||||||||| | |||||||  |||   ||||||||||||||   |||
Db     174784 CCCGGGAGGCGAAGGTTGCAGTGAGCCGAGATCACACTATTGCACTCCAGCCTGTTGACA 174725

Qy        656 GAGCAAGACTCCAT------AAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTA 709
              |||  |||||||||      ||| | || |||||     ||| ||  ||||| |||  |
Db     174724 GAGTGAGACTCCATGTCAAAAAAAAAAAAAAAAGAAAGAAAACTGCATAAATTAGTGTTT 174665

Qy        710 CCTATCTTCA-TTTAAGAAATTCATCTTTGTTCATTTATTTTTACT 754
              || |   || ||| |   |||||| | ||||  ||||  | | ||
Db     174664 TCTGTGGGCATTTTCATTCATTCATTTGTGTTGTTTTACATATGCT 174619
```

RESULT 42
HSXDPA/c
LOCUS HSXDPA 305000 bp DNA linear PRI 29-MAY-2003
DEFINITION Homo sapiens chromosome X sequence from 6 PACs 1 BAC and 1 cosmid, region GJB1-DXS559 map Xq13.1; segment 1/2, complete sequence.
ACCESSION AL590762 AJ239319
VERSION AL590762.1 GI:13751769
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ramser,J., Heitmann,K., Sudbrak,R., Kosiura,A., Klages,S., Steffens,C., Borzym,K., Kube,M., Lehrack,S., Marquardt,I., Schuelzchen,S., Starke,A., Thompson,C., Hennig,S., Francis,F., Nemeth,A., Monaco,A., Lehrach,H. and Reinhardt,R.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 305000)
AUTHORS MPIMG.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-1999) MPIMG, Abt.Lehrach, Max Planck Institut Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany
COMMENT On May 14, 2001 this sequence version replaced gi:11228442.
Bases 300001. .305000 overlap with AL590763 (HSXDPB).
FEATURES Location/Qualifiers
source 1. .305000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq13.1"
/clone="PAC RPCI-1 56F19"
/clone_lib="RPCI1,3-5 Human PAC library, originating institute: Roswell Park Cancer Institute, Creator: Pieter de Jong, P. Ioannou"
/note="region between markers GJB1-DXS559"
ORIGIN

Query Match 10.9%; Score 218.8; DB 8; Length 305000;
Best Local Similarity 62.0%; Pred. No. 1.7e-35;
Matches 424; Conservative 0; Mismatches 237; Indels 23; Gaps 4;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||||||  |||  ||||||||||||||||||||||||||||||||||||||||||
Db     117592 TGATCCACCCACCTCTGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCTG 117533

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  ||     | |||   ||| | | ||          |  ||      |  || |
Db     117532 GAC-TATAAAACTGTTAGAAGAAAACATAAAGATAAATCTTCATGACCTTGAATTTGGCA 117474

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              |   ||  |  |   ||  ||||| ||    || |    |       || || |||| |
Db     117473 ATGAATTCTTAAATATAACACCCAAAGCAAAGACATCAAAAGAAAAAATAGATAACTTGG 117414

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |   |||                 | | |       |   | || | | |  | |
Db     117413 ACCACATCA---------------AAATGAAAACTTTTGGGGGGGGTGTCCTAAAAAAAT 117369

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |||| || | |  || |   ||   ||   | |||| |  |||||||||  |
Db     117368 TAAAACTTTTA----TGCTTCAAATTACATTATTAAGAAAATGAAAATTTCAACACTGAA 117313

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              | |    | |    | | |  || | ||      ||  ||  |   |  | ||
Db     117312 TGGGAAAAAATATTTGTAAATCATACATCTGATAAGGGGCTTGTATCCAGGATATACAAA 117253

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              ||  |    || | | |||   |      ||||   |  |||   || |||| |||||||
Db     117252 GAACTCTTACAACTCAATAATAAAAAGACAAATAAACTGGGCCTGGTGGCTCATGCCTGT 117193

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||||| |||||||||||||||||||||||||  ||||
Db     117192 AATCCCAGCACTTTGGGAGGCCAAGGCGGGCGGATCACCTGAGGTCAGGAGTTCGAGACT 117133

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||||||||||||||||||| |||||||||||||| |||||||||||| ||  ||
Db     117132 AGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAACACAAAAATTAGCCGGGCATG 117073

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||| |||||||||||||||| |||||  |||||  ||||| |||||||| |||  ||||
Db     117072 GTGGCGTGCACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTCAAACC 117013

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGA 657
              | |||||||||||||| || | || ||||    | || ||||||||||||||     ||
Db     117012 TGGGAGGTGGAGGTTGCGGTGAGCCAAGATTGCGCCATTGCACTCCAGCCTGGGTGACGA 116953

Qy        658 GCAAGACTCCATAAAGACAACAAA 681
              |  | |||||||    | || |||
Db     116952 GTGAAACTCCATCTCAAAAAAAAA 116929
```

RESULT 43
HS568C11
LOCUS HS568C11 103681 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP4-568C11 on chromosome 20p11.21-11.23 Contains the CST7 gene for cystatin F (leukocystatin)(cystatin 7, leukocystatin, cystatin-like metastasis-associated protein, CMAP), the gene C20ORF3 for chromosome 20 open reading frame 3, the 3' end of the ACAS2L gene for acetyl-Coenzyme A synthetase (AMP forming)-like and a CpG island, complete sequence.
ACCESSION AL035661
VERSION AL035661.16 GI:6015535
KEYWORDS HTG; ACAS2L; AMP-binding; C20ORF3; CMAP; CpG island; CST7; cystatin; leukocystatin.
SOURCE Homo sapiens (human)

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 103681)
  AUTHORS   Barlow,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 7, 1999 this sequence version replaced gi:6002138.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            RP4-568C11 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .103681
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="p11.21-11.23"
                     /clone="RP4-568C11"
                     /clone_lib="RPCI-4"
     misc_feature    1
                     /note="Clone_left_end: RP4-568C11"
     gene            25136. .35834
                     /gene="CST7"
                     /locus_tag="RP4-568C11.3-001"
     mRNA            join(25136. .25481,33193. .33365,34834. .34950,35541. .35834)
                     /gene="CST7"
                     /locus_tag="RP4-568C11.3-001"
                     /note="match: ESTs: AA089317 AA089339 AI386371 AI707662
                     AI734965 AI927615 AL559378 AL559379 AW467354 AW514304
                     BE686225 BF523613 BF898503 BG535774 BG547285 BI768143
                     BI837059 BM253995 BM363008 BM363306 BM363506 BM363782
                     BM363821 BM365975 BM561663 BM564264 BM564523 BM917046
                     BQ051538 BQ051619 BQ052118 BQ053137 BQ053396 BQ053551
                     BQ054014 BQ054230 N47763
                     match: cDNAs: AB015225 AF031824 AF036342 BC015507"
     CDS             join(25346. .25481,33193. .33365,34834. .34950,35541. .35618)
                     /gene="CST7"
                     /locus_tag="RP4-568C11.3-001"
                     /standard_name="OTTHUMP00000030465"
                     /note="match: proteins: O76096 Q9UED4"
                     /codon_start=1
                     /protein_id="CAB75498.2"
                     /db_xref="GI:56203089"
                     /db_xref="Genew:2479"
                     /db_xref="GOA:O76096"
                     /db_xref="HSSP:1G96"
                     /db_xref="InterPro:IPR000010"
                     /db_xref="InterPro:IPR003243"
                     /db_xref="UniProt/Swiss-Prot:O76096"
                     /translation="MLPEKALHGHPQLPRTVPTRAAMRAAGTLLAFCCLVLSTTGGPS
                     PDTCSQDLNSRVKPGFPKTIKTNDPGVLQAARYSVEKFNNCTNDMFLFKESRITRALV
                     QIVKGLKYMLEVEIGRTTCKKNQHLRLDDCDFQTNHTLKQTLSCYSEVWVVPWLQHFE
                     VPVLRCH"
     misc_feature    32890. .33079
                     /gene="CST7"
                     /locus_tag="RP4-568C11.3-001"
                     /note="Tandem repeat. Region contains single clone and
                     forced join. Assembly is consistent with the digest."
     polyA_signal    35803. .35808
                     /gene="CST7"
                     /locus_tag="RP4-568C11.3-001"
     polyA_site      35834
                     /gene="CST7"
                     /locus_tag="RP4-568C11.3-001"
     gene            complement(38831. .68657)
                     /gene="C20orf3"
                     /locus_tag="RP4-568C11.2-001"
     mRNA            complement(join(38831. .39928,44798. .44990,45432. .45566,
                     46103. .46277,47366. .47482,49551. .49643,54673. .54788,
                     59809. .59925,68500. .68657))
                     /gene="C20orf3"
                     /locus_tag="RP4-568C11.2-001"
                     /note="match: ESTs: AA023064 AA057573 AA078933 AA079231
                     AA128202 AA130982 AA164222 AA172232 AA179409 AA179549
                     AA190882 AA193559 AA199685 AA256230 AA287697 AA295645
                     AA295719 AA299550 AA299830 AA308436 AA335036 AA345847
                     AA350163 AA374277 AA393476 AA455403 AA464128 AA478826
                     AA479994 AA552922 AA602522 AA604095 AA632605 AA709451
                     AA737976 AA774045 AA775837 AA782787 AA810142 AA853299
                     AA861665 AA892916 AA908661 AA932982 AA974843 AA975875
                     AA989411 AI032386 AI042188 AI125617 AI140761 AI199766
                     AI200324 AI205248 AI208259 AI278121 AI301561 AI435846
                     AI470112 AI500350 AI526585 AI560081 AI750684 AI750995
                     AI750996 AI754666 AI754907 AI811737 AI819701 AI826600
                     AI828005 AI884917 AI961461 AI991422 AW005061 AW023303
                     AW028593 D20431 F12233 H12669 H26812 H26824 H27474 H28147
                     H47768 H82670 H94409 N32314 N40320 N41442 N57409 R11259
                     R12093 R13491 R18192 R37399 R42358 R54498 R55919 R56192
                     R71351 R79866 R83725 R83729 T03538 T23653 T23663 T30196
                     T34150 T34847 T48031 T55377 T55805 T66394 T73292 T73306
                     T73353 T73370 T77966 T84396 T93282 W63689 Z19171 Z28429
                     Z43384 Z44002
                     match: cDNAs: M57547"
     polyA_site      complement(38831)
                     /gene="C20orf3"
                     /locus_tag="RP4-568C11.2-001"
     polyA_site      complement(38850)
                     /gene="C20orf3"
                     /locus_tag="RP4-568C11.2-001"
     polyA_signal    complement(38864. .38869)
                     /gene="C20orf3"
                     /locus_tag="RP4-568C11.2-001"
     CDS             complement(join(39719. .39928,44798. .44990,45432. .45566,
                     46103. .46277,47366. .47482,49551. .49643,54673. .54788,
                     59809. .59925,68500. .68594))
                     /gene="C20orf3"
                     /locus_tag="RP4-568C11.2-001"
                     /standard_name="OTTHUMP00000030464"
                     /note="match: proteins: CE06440 CE16156 O81484 P15324
                     P18417 Q24160 Q9ZVB7"
                     /codon_start=1
                     /protein_id="CAB75499.1"
                     /db_xref="GI:6996428"
                     /db_xref="Genew:13238"
                     /db_xref="GOA:Q9HDC9"
                     /db_xref="InterPro:IPR004141"
                     /db_xref="UniProt/Swiss-Prot:Q9HDC9"
                     /translation="MSEADGLRQRRPLRPQVVTDDDGQAPEAKDGSSFSGRVFRVTFL
                     MLAVSLTVPLLGAMMLLESPIDPQPLSFKEPPLLLGVLHPNTKLRQAERLFENQLVGP
```

```
                     ESIAHIGDVMFTGTADGRVVKLENGEIETIARFGSGPCKTRDDEPVCGRPLGIRAGPN
                     GTLFVADAYKGLFEVNPWKREVKLLLSSETPIEGKNMSFVNDLTVTQDGRKIYFTDSS
                     SKWQRRDYLLLVMEGTDDGRLLEYDTVTREVKVLLDQLRFPNGVQLSPAEDFVLVAET
                     TMARIRRVYVSGLMKGGADLFVENMPGFPDNIRPSSSGGYWVGMSTIRPNPGFSMLDF
                     LSERPWIKRMIFKLFSQETVMKFVPRYSLVLELSDSGAFRRSLHDPDGLVATYISEVH
                     EHDGHLYLGSFRSPFLCRLSLQAV"
     gene            join(complement(AL080312.14:30098. .30921),
                     complement(AL080312.14:20414. .20510),
                     complement(AL080312.14:3088. .3287),
                     complement(99372. .99547),complement(98846. .98998),
                     complement(97295. .97442),complement(95916. .96053),
                     complement(91044. .91136),complement(89867. .89979),
                     complement(89442. .89568),complement(88718. .88845),
                     complement(88507. .88570),complement(85176. .85294),
                     complement(82138. .83847))
                     /locus_tag="RP4-568C11.1-001"
     mRNA            join(complement(AL080312.14:30098. .30921),
                     complement(AL080312.14:20414. .20510),
                     complement(AL080312.14:3088. .3287),
                     complement(99372. .99547),complement(98846. .98998),
                     complement(97295. .97442),complement(95916. .96053),
                     complement(91044. .91136),complement(89867. .89979),
                     complement(89442. .89568),complement(88718. .88845),
                     complement(88507. .88570),complement(85176. .85294),
                     complement(82138. .83847))
                     /locus_tag="RP4-568C11.1-001"
                     /note="match: ESTs: AI198432.1 AI356425.1 AI385955.1
                     AI478819.1 AI762466.1 AI983417.1 AW007194.1 AW024994.1
                     AW513684.1 BE870732.1 BF027291.1 BF116729.1 BF195060.1
                     BF514433.1 BG746986.1 BG758180.1 BG766356.1 BG866642.1
                     BI009737.1 BI548978.1 BM150326.1 BM680382.1 BQ187659.1
                     BQ364130.1 BQ366261.1 BQ641734.1
                     match: cDNAs: AB058749.1 AK027817.1 AL832939.1 BC032076.1
                     BC039261.1"
     gene            complement(join(82138. .83847,85176. .85905))
                     /locus_tag="RP4-568C11.1-003"
     mRNA            complement(join(82138. .83847,85176. .85905))
                     /locus_tag="RP4-568C11.1-003"
                     /note="match: ESTs: AA580377.1 AI198432.1 AI478819.1
                     BE703975.1 BE767662.1 BG758635.1 BM680382.1
                     match: cDNAs: AK090851.1 AL833041.1"

  Query Match            10.9%;  Score 218.4;  DB 8;  Length 103681;
  Best Local Similarity   81.1%;  Pred. No. 2.6e-35;
  Matches  279;  Conservative    0;  Mismatches   61;  Indels    4;  Gaps    2;

Qy        343 TAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGC 402
              ||| |||| |  ||||  | | |  |||||  ||| | |  || |||||  ||||||||
Db      80653 TAACAAAAACTATAAAATAACAAAAACTAAATGTTTAGAAATTTTAGAAATAGGCCAGGC 80712

Qy        403 GCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGA 462
              |  || ||||   |||||||||||||||||||||||||||||| ||| | |||||||.|
Db      80713 GTGGTGGCTCACACCTGTAATCCCAACACTTTGGGAGGCCAAGGCGGGTGGTATCACCTAA 80772

Qy        463 GGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGC 522
              |||||||||||| |||||||||||||||||||| | |||||  |||||||||||||  |
Db      80773 GGTCAGGAGTTTGAGACCAGCCTGGCCAACATGTTAAAACCTCGTCTCTACTAAAAATAC 80832

Qy        523 AAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGG 582
              |||||||||||||||||||||| |||||||||||||||| |||||| ||||| |||||
Db      80833 AAAAATTAGCCAGGTGTGGTGGCGTGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGC 80892

Qy        583 AGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCA 639
              |||||||||||||| |   |||||||||||||||||| | | ||||   || |||| ||
Db      80893 AGGAGAATTGCTTGACCTGGGGAGGTGGAGGTTGCAGTGAGCTGAGATGGTGCCACTACA 80952

Qy        640 CTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||  || ||  |||||||||||||||||    | || ||||
Db      80953 CTCCAAACTGGGTGACAGAGCAAGACTCCATCTCAAAAAAAAAA 80996


RESULT 44
AC020634/c
```

```
LOCUS       AC020634              181174 bp    DNA     linear   PRI 29-JUN-2002
DEFINITION  Homo sapiens 3 BAC RP11-9N20 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC020634
VERSION     AC020634.16  GI:19033427
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 181174)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181174)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 181174)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 181174)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   5  (bases 1 to 181174)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 1, 2002 this sequence version replaced gi:18652479.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .181174
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-9N20"
     misc_feature    complement(1. .2004)
                     /note="overlaps bases 1. .2004 of clone AC117381"
                     /function="clone overlap"
     repeat_region   complement(188. .560)
                     /rpt_family="MLT1A1"
     repeat_region   3310. .3337
                     /rpt_family="AT_rich"
     repeat_region   complement(3338. .5322)
                     /rpt_family="L1MA5A"
     repeat_region   5323. .5377
                     /rpt_family="(TA)n"
     repeat_region   complement(5378. .5389)
                     /rpt_family="L1MA5A"
     repeat_region   complement(5386. .5853)
                     /rpt_family="L1M2"
     repeat_region   complement(5846. .6339)
                     /rpt_family="L1M2"
     repeat_region   complement(7364. .7707)
                     /rpt_family="L1M2"
     repeat_region   complement(7904. .8205)
                     /rpt_family="AluJo"
     repeat_region   complement(8597. .8682)
                     /rpt_family="L2"
     repeat_region   8683. .8990
                     /rpt_family="AluSq"
     repeat_region   complement(8991. .9981)
                     /rpt_family="L2"
     repeat_region   complement(10829. .11216)
                     /rpt_family="MSTD"
     repeat_region   11899. .11925
                     /rpt_family="AT_rich"
     repeat_region   12044. .12346
                     /rpt_family="AluSx"
     repeat_region   12356. .12460
                     /rpt_family="(GGAA)n"
     repeat_region   13263. .13284
                     /rpt_family="(A)n"
     repeat_region   complement(13499. .13940)
                     /rpt_family="LTR40b"
     repeat_region   complement(13966. .14096)
                     /rpt_family="MIR"
     repeat_region   complement(14179. .14317)
                     /rpt_family="MIR"
     repeat_region   complement(14366. .14643)
                     /rpt_family="AluJb"
     repeat_region   complement(16681. .16824)
                     /rpt_family="MIR"
     repeat_region   17093. .17211
                     /rpt_family="L2"
     repeat_region   17980. .18260
                     /rpt_family="AluSx"
     repeat_region   18261. .18288
                     /rpt_family="(GAAA)n"
     repeat_region   18471. .18501
                     /rpt_family="AT_rich"
     repeat_region   complement(19099. .19176)
                     /rpt_family="MER94"
     repeat_region   19264. .19555
                     /rpt_family="AluSx"
     STS             20829. .21105
                     /standard_name="170359"
     repeat_region   complement(23518. .23829)
                     /rpt_family="AluSq"
     repeat_region   24089. .24348
                     /rpt_family="MIR"
     repeat_region   24297. .24359
                     /rpt_family="L2"
     repeat_region   complement(25626. .25925)
                     /rpt_family="AluSq"
     repeat_region   25976. .26107
                     /rpt_family="MIR"
     repeat_region   26108. .26395
                     /rpt_family="AluSx"
     repeat_region   26396. .26417
                     /rpt_family="(TAAA)n"
```

```
Query Match             10.9%;  Score 218.4;  DB 8;  Length 181174;
Best Local Similarity   74.1%;  Pred. No. 2.3e-35;
Matches  304;  Conservative    0;  Mismatches  101;  Indels    5;  Gaps    2;

Qy       287 ATTTCAACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAAT 346
             |||  |  | | | |     | |||||  |  |  |  | |||  |  |    |   |
Db     47159 ATTGGACAATGCAATTCTGTATCTTGAACAGCTCACTCTCATTGAAGAAGCAAAAGATAA 47100

Qy       347 AAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAG 406
             | |||  |  |    | || |  |  ||||  |||  | |    | | |||| ||| |||
Db     47099 ACAATAATGGATCAAATTATAAATTGATTAAAAAACTGATAAAGAGTAGGCCGGGCACAG 47040

Qy       407 TAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTC 466
             | ||||  ||||||||||||| |||||||||||||| ||||||||||||||||||||||||
Db     47039 TGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCACCTGAGGTC 46980

Qy       467 AGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAA 526
             |||||||  |||||||||| ||||||||||||||||||||||||||||||||  |||||
```

```
Db      46979 AGGAGTTCGAGACCAGCCTAGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAA 46920

Qy        527 ATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGA 586
              ||||||| |||||||||||| || |||||||||||| |||||||||||| ||||| ||||
Db      46919 ATTAGCCGGGTGTGGTGGCACGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGA 46860

Qy        587 GAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCC 643
              |  | ||| |||||| ||||| |||||||||||| | ||| ||   ||||||||||||||
Db      46859 GTGTCGCTCGAACCTGGGAGGCGGAGGTTGCAGTGAACCGCGATTGTGTCACTGCACTCC 46800

Qy        644 AGCCT--GGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAA 691
              ||| |  |||||||||| ||||||||||     | || ||||  ||   ||
Db      46799 AGCTTGGGGCAACAGAGTAAGACTCCATCTCAAAAAAAAAAAATTGATAA 46750


RESULT 45
AC161061/c
LOCUS       AC161061               187825 bp    DNA     linear   HTG 11-MAY-2005
DEFINITION  Pan troglodytes chromosome UNKNOWN clone CH251-489A6, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC161061
VERSION     AC161061.1  GI:63055279
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 187825)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Pan troglodytes clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 187825)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2005) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 187825)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-2005) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: C_AB0489A06
            -----------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 181623 bases at least Q40
            Consensus quality: 182527 bases at least Q30
            Consensus quality: 183339 bases at least Q20
            -----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1599: contig of 1599 bp in length
            *     1600     1699: gap of unknown length
            *     1700     3234: contig of 1535 bp in length
            *     3235     3334: gap of unknown length
            *     3335     4681: contig of 1347 bp in length
            *     4682     4781: gap of unknown length
            *     4782     7438: contig of 2657 bp in length
            *     7439     7538: gap of unknown length
            *     7539    21710: contig of 14172 bp in length
            *    21711    21810: gap of unknown length
            *    21811    38270: contig of 16460 bp in length
            *    38271    38370: gap of unknown length
            *    38371    73523: contig of 35153 bp in length
            *    73524    73623: gap of unknown length
            *    73624   115929: contig of 42306 bp in length
            *   115930   116029: gap of unknown length
            *   116030   187825: contig of 71796 bp in length.
FEATURES             Location/Qualifiers
     source          1. .187825
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="UNKNOWN"
                     /clone="CH251-489A6"
     misc_feature    1. .1599
                     /note="assembly_name:Contig550"
     gap             1600. .1699
                     /estimated_length=unknown
     misc_feature    1700. .3234
                     /note="assembly_name:Contig572"
     gap             3235. .3334
                     /estimated_length=unknown
     misc_feature    3335. .4681
                     /note="assembly_name:Contig584"
     gap             4682. .4781
                     /estimated_length=unknown
     misc_feature    4782. .7438
                     /note="assembly_name:Contig587"
     gap             7439. .7538
                     /estimated_length=unknown
     misc_feature    7539. .21710
                     /note="assembly_name:Contig588
                     clone_end:SP6
                     vector_side:right"
     gap             21711. .21810
                     /estimated_length=unknown
     misc_feature    21811. .38270
                     /note="assembly_name:Contig589"
     gap             38271. .38370
                     /estimated_length=unknown
     misc_feature    38371. .73523
                     /note="assembly_name:Contig590"
     gap             73524. .73623
                     /estimated_length=unknown
     misc_feature    73624. .115929
                     /note="assembly_name:Contig591"
     gap             115930. .116029
                     /estimated_length=unknown
     misc_feature    116030. .187825
                     /note="assembly_name:Contig592"
ORIGIN

  Query Match            10.9%;  Score 218.4;  DB 14;  Length 187825;
  Best Local Similarity   59.1%;  Pred. No. 2.3e-35;
  Matches  447;  Conservative    0;  Mismatches  301;  Indels    8;  Gaps    4;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||| || |||| |||||||| |||||||||||||||| |||||||||||| |||| |
Db      16226 TGATCCGCCCGCCTCGGCCTCCCTAAGTGCTGGGATTACAAGCGTGAGCCACCGCGCCAG 16167

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  | |  |  || |||  | |  ||    |  | |  |    | |  |      |
Db      16166 GCC--AGGGTTGAATAATTTTTAAATGCATTTTACACTTGGAGAAGATTGTACCAAAATC 16109

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | | ||     |||      |     || |  |   |   |  |   || |
Db      16108 ATCCATTTGCATTTTCAGTGAATGACATATAAACATTTTAGTCATTATATATAAGCTAAA 16049
```

```
Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | | |    | ||| || || | |   |   |   |   |   |||    ||
Db      16048 AGTATAAAAAGATTGTTTATGTATTTTTTTAGCATATCTACTTTTCATTTGTGAGAACCC 15989

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              || |   | |    |   | | |           ||||    | || ||  | ||
Db      15988 TAGAGGAAAAAGTGTTACAAAAAAGAAAACACCCCCCAAAATCTCTACTTTAATGAATAA 15929

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
               |||| | || ||  | |    ||||  |||   |   | |||     ||     |
Db      15928 AAGCTAATGATGTAGGCTGCATTCCTATTATTTAGTTCCTCTAAATGTCTATTCCCTTAT 15869

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              |||||    |   |||||||     |      | |||| |||   || |||    |||||
Db      15868 TAATAACTCTGCTTTATTAAAATAACTGT--CTCGGCCGGGCATGGTGGCTGACACCTGT 15811

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| ||||||||||||  |||||||||||||||||||| |||||| |||| |
Db      15810 AATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACCTGAGGTCGGGAGTTCAAGAGC 15751

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||| |||||||||||||||| ||| |||||||||| |||||||||  ||  ||
Db      15750 AGCCTGGCTAACATGGTGAAACCCCGTCACTACTAAAAACACAAAAATTAGCTGGGCATG 15691

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              || |   ||||||||| | ||| |||||| ||||| ||||| ||||||||||||||||||
Db      15690 GTCGTGGGCACCTGTAATTCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACC 15631

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACAG 656
               |||||||||||||||||| | || ||||   | ||||||||||||||| || |||||
Db      15630 CAGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCGCCACTGCACTCCAGCCTGGGTAACAG 15571

Qy        657 AGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCT 716
              | | ||||||||||  | | || ||||       |||   | |   |  | |    ||| |
Db      15570 ACCGAGACTCCATCTAAAAAAAAAAAAAACAAAAAAAAATCTTTTTTTGGGGGGGTTATAT 15511

Qy        717 TCATTTAAGAAATTCATCTTTGTTCATTTATTTTTA 752
                |||   | |  |   | |  || ||  | ||
Db      15510 GACCCTAAATTAGTGCTAAATATATATCTAAATATA 15475
```

Search completed: January 22, 2006, 13:37:24
Job time : 6828.68 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2006, 09:21:22 ; Search time 805.081 Seconds
(without alignments)
16564.866 Million cell updates/sec

Title: US-09-728-552A-3_COPY_39000_41000
Perfect score: 2001
Sequence: 1 tgatccaccagccttggcct.....ctaattcagtgaacacttca 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2001	100.0	80595	2	AAV83939	Aav83939 HC-contig
2	1959.4	97.9	80240	2	AAV83940	Aav83940 NC-contig
3	221.2	11.1	121162	3	AAC66548	Aac66548 Human kin
4	215.6	10.8	153170	12	ADQ17382	Adq17382 Human sof
c 5	214.4	10.7	325791	4	AAS43104	Aas43104 Human Oes
6	213.6	10.7	2013	6	ACC57389	Acc57389 Human mac
c 7	213.2	10.7	11173	9	ABZ70982	Abz70982 Human HKR
8	212.2	10.6	59725	11	ACN44390	Acn44390 Human gen
c 9	211.2	10.6	108316	10	ADC87336	Adc87336 Human GPC
c 10	210.8	10.5	11950	4	AAS41753	Aas41753 Genomic s
c 11	210.8	10.5	11950	4	ABA06813	Aba06813 Human gen
c 12	210.8	10.5	11950	5	AAS29743	Aas29743 Human end
c 13	210.8	10.5	11950	6	ABV84150	Abv84150 Human pol
c 14	210.6	10.5	128034	10	ADE43582	Ade43582 Polymorph
c 15	210.6	10.5	128034	10	ADE43581	Ade43581 Human IDE
c 16	210.6	10.5	128034	12	ADH54059	Adh54059 Human IDE
c 17	210.6	10.5	128034	12	ADH54060	Adh54060 Human IDE
18	210.6	10.5	202100	10	ADE43315	Ade43315 Human IDE
19	210.6	10.5	202100	12	ADH54357	Adh54357 Human IDE
c 20	210.2	10.5	110000	14	AEA61169_2	Continuation (3 of
c 21	210.2	10.5	203654	10	ABX16034	Abx16034 Human gen
22	210.2	10.5	260160	12	ADQ20017	Adq20017 Human sof
c 23	209.2	10.5	178024	12	ADQ97721	Adq97721 Human can
c 24	208.8	10.4	93390	10	ADD71350	Add71350 Glutamine
25	208.8	10.4	243934	14	ADZ13446	Adz13446 Human can
c 26	208.6	10.4	17588	10	ADC86824	Adc86824 Human GPC
c 27	208.6	10.4	63266	12	ADO48536	Ado48536 Human neu
c 28	208.6	10.4	63266	14	ADX80721	Adx80721 Human nid
29	208.6	10.4	101685	12	ADQ17329	Adq17329 Human sof
30	208.4	10.4	8553	4	AAL04768	Aal04768 Human rep
31	208.4	10.4	8553	4	ABL97671	Abl97671 Human tes
32	208.4	10.4	51928	11	ACN44462	Acn44462 Human gen
33	208.4	10.4	52124	14	ADZ12979	Adz12979 Human can
c 34	208.2	10.4	27154	4	AAL05708	Aal05708 Human rep
35	208.2	10.4	254396	12	ADQ97557	Adq97557 Human can
36	208	10.4	32145	4	AAK68575	Aak68575 Human imm
37	208	10.4	32145	4	AAK68491	Aak68491 Human imm
38	208	10.4	83450	10	AAD58977	Aad58977 Human pho
39	208	10.4	83450	10	ADG88582	Adg88582 Human pho
40	208	10.4	83450	14	AEB47870	Aeb47870 Human nov
c 41	208	10.4	110000	12	ADQ97960_0	Adq97960 Human can
c 42	207	10.3	101209	11	ACN44154	Acn44154 Human gen
43	207	10.3	243390	13	ABD33366	Abd33366 Human can
c 44	206.8	10.3	22700	4	AAK70746	Aak70746 Human imm
45	206.8	10.3	59365	13	ABD32596	Abd32596 Human can
46	206.8	10.3	326002	13	ABD32843	Abd32843 Human can
47	206.6	10.3	120447	12	ADQ97391	Adq97391 Human can
48	206.6	10.3	220895	6	ABK84798	Abk84798 Human cDN
49	206.6	10.3	220895	13	ADR52737	Adr52737 Drug ther
50	206.4	10.3	1727	4	AAH17438	Aah17438 Human cDN
c 51	206.2	10.3	606	6	ABN62170	Abn62170 Human can
c 52	206.2	10.3	41991	13	ABD32684	Abd32684 Human can
c 53	206.2	10.3	81001	4	AAF30035	Aaf30035 Human apo
54	206.2	10.3	188971	12	ADL08108	Adl08108 Human gen
c 55	206	10.3	4900	6	AAL43026	Aal43026 Human P2X
56	206	10.3	25837	4	AAK85952	Aak85952 Human imm
c 57	206	10.3	36221	4	AAS00624	Aas00624 Human dea
c 58	206	10.3	42123	11	ACN44906	Acn44906 Human gen
c 59	206	10.3	56580	13	ADT25932	Adt25932 Wild type
c 60	206	10.3	56580	13	ADU23759	Adu23759 Human ATP
61	206	10.3	147708	6	ABQ88154	Abq88154 Human ost
62	206	10.3	197526	13	ABD33389	Abd33389 Human can
c 63	205.8	10.3	24333	12	ADJ12497	Adj12497 DNA fragm
c 64	205.8	10.3	24333	12	ADJ12606	Adj12606 DNA fragm
c 65	205.8	10.3	72860	13	ACN37232	Acn37232 Human per
c 66	205.8	10.3	96594	10	ADC85476	Adc85476 Human Msf
c 67	205.8	10.3	96595	9	ADA02996	Ada02996 Human Msf
c 68	205.8	10.3	96595	10	ADB72734	Adb72734 Human Msf
c 69	205.8	10.3	96595	12	ADM74591	Adm74591 Human car
c 70	205.6	10.3	3226	11	ADM02038	Adm02038 Human cDN
71	205.6	10.3	191127	10	ADH13788	Adh13788 Prostate
c 72	205.4	10.3	544	4	AAH10191	Aah10191 Human cDN
c 73	205.4	10.3	44196	4	AAK77458	Aak77458 Human imm
74	205.4	10.3	76180	13	ABD33385	Abd33385 Human can
75	205.4	10.3	104083	11	ACN44294	Acn44294 Human gen
76	205.4	10.3	105223	14	ADZ12953	Adz12953 Human can
c 77	205.2	10.3	87847	13	ABD33066	Abd33066 Human can
c 78	205.2	10.3	112241	13	ABD33491	Abd33491 Human can
c 79	205.2	10.3	201143	6	ABK83568	Abk83568 Human DNA
80	205	10.2	2844	4	AAK91523	Aak91523 Human dig
81	205	10.2	2844	5	AAS32177	Aas32177 Human liv
82	205	10.2	2844	6	ABN90532	Abn90532 Human liv
83	205	10.2	2844	11	ADJ15445	Adj15445 Human liv
84	205	10.2	34668	10	ACC00044	Acc00044 Nucleotid
85	205	10.2	276820	11	ADP75188	Adp75188 Human ADA
86	204.8	10.2	2024	6	ABZ82521	Abz82521 Human sec
87	204.8	10.2	24167	5	ABA16132	Aba16132 Human ner
88	204.8	10.2	32152	4	AAK89020	Aak89020 Human dig
c 89	204.8	10.2	32152	4	AAK91534	Aak91534 Human dig
c 90	204.8	10.2	32152	4	AAI57791	Aai57791 Human col
91	204.8	10.2	32152	5	AAS39621	Aas39621 Genomic s
c 92	204.8	10.2	32152	6	ABS99968	Abs99968 Genomic D

	93	204.8	10.2	32152	9	ADB32581	Adb32581 Human nov
c	94	204.8	10.2	32152	10	ADB93121	Adb93121 Human col
	95	204.8	10.2	32204	4	AAK89019	Aak89019 Human dig
c	96	204.8	10.2	32204	4	AAK91533	Aak91533 Human dig
c	97	204.8	10.2	32204	4	AAI57790	Aai57790 Human col
	98	204.8	10.2	32204	5	AAS39620	Aas39620 Genomic s
c	99	204.8	10.2	32204	6	ABS99967	Abs99967 Genomic D
	100	204.8	10.2	32204	9	ADB32580	Adb32580 Human nov

ALIGNMENTS

RESULT 1
AAV83939
ID AAV83939 standard; DNA; 80595 BP.
XX
AC AAV83939;
XX
DT 03-MAR-1999 (first entry)
XX
DE HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 8; Fig 6; 540pp; English.
XX
CC The present sequence represents the HC-contig derived from normal human
CC chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC produce an unusual chromosomal marker referred to as mardel(10). The
CC mardel(10) marker is mitotically stable and contains a functional
CC neocentromere at a location regarded as non-centromeric. This
CC neocentromere maps to q25.2 on chromosome 10. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in a
CC compatible cell, of replicating, acting as extra-chromosomal element and
CC segregating during cell division. The sequences can be used to construct
CC artificial chromosomes for use in gene therapy comprising a replicable,
CC segregating nucleic acid that confers a specific phenotype on cells.
CC Human artificial chromosomes can propagate in human cells and carry large
CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC they are not mutagenic. The artificial chromosomes are also useful for
CC generation of transgenic plants and animals, in production of proteins
CC and to make diagnostic reagents, e.g. for expression of cytokines,
CC receptors and growth factors, or to increase the copy number of a gene in
CC a cell. The constructs may also be used for functional and structural
CC analysis of chromosomes
XX

SQ Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

Query Match 100.0%; Score 2001; DB 2; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39000 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 39059

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39060 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 39119

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39120 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 39179

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39180 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 39239

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39240 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 39299

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39300 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39359

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39360 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39419

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39420 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39479

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39480 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39539

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39540 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39599

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39600 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39659

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39660 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39719

Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39720 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39779

Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39780 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39839

Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39840 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39899

Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39900 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 39959

Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39960 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 40019
```

```
Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40020 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 40079

Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40080 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 40139

Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40140 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 40199

Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40200 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 40259

Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40260 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 40319

Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40320 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40379

Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40380 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40439

Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40440 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40499

Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40500 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 40559

Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40560 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 40619

Qy       1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40620 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 40679

Qy       1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40680 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40739

Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40740 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40799

Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40800 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40859

Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40860 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 40919

Qy       1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40920 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 40979

Qy       1981 TCTAATTCAGTGAACACTTCA 2001
             |||||||||||||||||||||
Db      40980 TCTAATTCAGTGAACACTTCA 41000
```

```
RESULT 2
AAV83940
ID   AAV83940 standard; DNA; 80240 BP.
XX
AC   AAV83940;
XX
DT   03-MAR-1999  (first entry)
XX
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 9; Fig 16A; 540pp; English.
XX
CC   The present sequence represents the NC-contig derived from a mutated
CC   human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC   chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC   mitotically stable and contains a functional neocentromere at a location
CC   regarded as non-centromeric. This neocentromere maps to q25.2 on
CC   chromosome 10. The specification describes nucleic acid sequences derived
CC   from a eukaryotic chromosome, including a neocentromere or its functional
CC   derivative or hybrid, that are able, in a compatible cell, of
CC   replicating, acting as extra-chromosomal element and segregating during
CC   cell division. The sequences can be used to construct artificial
CC   chromosomes for use in gene therapy comprising a replicable, segregating
CC   nucleic acid that confers a specific phenotype on cells. Human artificial
CC   chromosomes can propagate in human cells and carry large amounts of DNA
CC   (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC   mutagenic. The artificial chromosomes are also useful for generation of
CC   transgenic plants and animals, in production of proteins and to make
CC   diagnostic reagents, e.g. for expression of cytokines, receptors and
CC   growth factors, or to increase the copy number of a gene in a cell. The
CC   constructs may also be used for functional and structural analysis of
CC   chromosomes
XX
SQ   Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

  Query Match             97.9%;  Score 1959.4;  DB 2;  Length 80240;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1988;  Conservative    0;  Mismatches   11;  Indels    2;  Gaps    2;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38738 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 38797

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38798 GTCGAATGTTTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 38857

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      38858 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 38917
Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
Db      38918 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGC-TTCAAGGAGGAA 38976
Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
Db      38977 TAAACCTATGAAAAGTGTTGAGGATAGTGTCTGATATGAAATAAGGA-TTCAACAAGTAG 39035
Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
Db      39036 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39095
Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
Db      39096 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39155
Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
Db      39156 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39215
Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
Db      39216 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39275
Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
Db      39276 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39335
Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
Db      39336 TAGGAGCTGGAGGTTGCAGTGACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39395
Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
Db      39396 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39455
Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
Db      39456 TTAAGAAATTCATCTTTGTTCATCTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39515
Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
Db      39516 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39575
Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
Db      39576 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39635
Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
Db      39636 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAGGAT 39695
Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
Db      39696 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 39755
Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
Db      39756 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 39815
Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
Db      39816 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 39875
Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
Db      39876 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 39935
Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
Db      39936 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 39995
```

```
Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
Db      39996 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 40055
Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
Db      40056 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40115
Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
Db      40116 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40175
Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
Db      40176 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40235
Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
Db      40236 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 40295
Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
Db      40296 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCAGAATGGAATCCTGTAGAATGTTCAC 40355
Qy       1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
Db      40356 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTAGTAAACTATC 40415
Qy       1681c TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
Db      40416 TTAACCAAGACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40475
Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
Db      40476 AATAATAAATCTGTCCAGTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40535
Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
Db      40536 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40595
Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
Db      40596 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 40655
Qy       1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
Db      40656 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 40715
Qy       1981 TCTAATTCAGTGAACACTTCA 2001
Db      40716 TCTAATTCAGTGAACACTTCA 40736
```

```
RESULT 3
AAC66548
ID   AAC66548 standard; DNA; 121162 BP.
XX
AC   AAC66548;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Human kinesin-like protein HKLP coding sequence contig SEQ ID NO: 1.
XX
KW   Human; kinesin-like protein; HKLP; KIF1; cell division; cancer;
KW   intracellular transport; neurological disorder; infertility;
KW   biallelic marker; spontaneous abortion; neonatal chromosome disorder;
KW   aneuploidy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200063375-A1.
XX
```

```
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-IB000562.
XX
PR   20-APR-1999;   99US-0130217P.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Dufaure-Gare I,  Grel P;
XX
DR   WPI; 2000-665242/64.
XX
PT   An isolated or purified human kinesin-like protein (HKLP) encoding
PT   polynucleotide used to detect HKLP polynucleotides in a sample comprises
PT   a contiguous span of at least 12 nucleotides.
XX
PS   Claim 1; Page 143-175; 199pp; English.
XX
CC   The present invention describes the coding and protein sequences of the
CC   human kinesin-like protein HKLP. It is thought that the protein could be
CC   involved in neurological disorders, infertility, spontaneous abortion,
CC   neonatal chromosome disorders, aneuploidy and cancers. This is due to its
CC   function in the movement of microtubules. The protein shows homology to
CC   the murine KIF1A and KIF1B proteins. The sequences disclosed in the
CC   invention can be used in the isolation of similar human proteins and in
CC   vector production. In addition, the biallelic markers shown can be used
CC   in disease diagnosis and population studies
XX
SQ   Sequence 121162 BP; 33272 A; 24108 C; 25842 G; 37919 T; 0 U; 21 Other;

  Query Match            11.1%;  Score 221.2;  DB 3;  Length 121162;
  Best Local Similarity  83.8%;  Pred. No. 2.2e-36;
  Matches  263;  Conservative    0;  Mismatches   48;  Indels    3;  Gaps    1;

Qy        388 TGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGT 447
              |||| | |||||||||||||| || ||  |||||||||||| ||||||||| |||| ||||
Db      25335 TGAAGTCGGCCAGGCGCAGTGGCCCCCACCTGTAATCCCAGCACTTTGGGTGGCCGAGGT 25394

Qy        448 GGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGT 507
              |||||||||||| |||||||||||||||| |||||||||||||||||||| ||||||||  |
Db      25395 GGGCGGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGATGAAACCCCAT 25454

Qy        508 CTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTAC 567
              |||||||||||  ||||||| || | || |||||| |||||||||||| ||| | ||||
Db      25455 CTCTACTAAAAATACAAAAATGAGTCGGGCGTGGTGCCATGCACCTGTAATCCTAGCTAC 25514

Qy        568 TCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGA 627
              |||||||| ||||| |||||||||||||||||| ||||| |||||||||||  | ||||
Db      25515 TCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGA 25574

Qy        628 GAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGC 684
              |||   | |||||||||||||||||  ||||||| |||||||||     | || ||||
Db      25575 GATGGCGCCACTGCACTCCAGCCTGGAGACAGAGCGAGACTCCATCTCAAAAAAAAAAA 25634

Qy        685 TTTGAAATTGTGTA 698
                  ||||| || |
Db      25635 AAATAAATTCTGAA 25648


RESULT 4
ADQ17382
ID   ADQ17382 standard; DNA; 153170 BP.
XX
AC   ADQ17382;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 199.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 199; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 153170 BP; 39941 A; 37835 C; 36509 G; 38885 T; 0 U; 0 Other;

  Query Match            10.8%;  Score 215.6;  DB 12;  Length 153170;
  Best Local Similarity  77.6%;  Pred. No. 3.4e-35;
  Matches  287;  Conservative    0;  Mismatches   79;  Indels    4;  Gaps    2;

Qy        311 TGAAGATTTAAGAGTTATTTATTACAA-CTATTTAATAAAATTTTAAAAACTAATACACT 369
              | || |||||| | ||| | |||| ||     || |  |  ||  | |    | ||||
Db      73952 TTAAAATTTAAAATTTACTCATTATAATGCCATTTTTCTAGATTACAGGAGACGTTCACT 74011

Qy        370 TAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAAC 429
              |     ||| | |  |  | |  |||| |||||||| |||| ||||||||||||||||
Db      74012 TTTCGCCTTATATAATTCAGTATAAGGCCGGGCGCAGTGGCTCATGCCTGTAATCCCAAA 74071

Qy        430 ACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCC 489
              |||||||||||||| ||| |||||||||||| |||||||||| ||  || |||||||| ||
Db      74072 ACTTTGGGAGGCCGAGGCGGGCGGATCACTTGAGGTCAGGAATTCGAGGCCAGCCTGACC 74131

Qy        490 AACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGC 549
              ||||||| ||||| | ||||||||||||||  ||||||||||||||| |||||||| ||
Db      74132 AACATGGCGAAACTCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGCACGC 74191

Qy        550 ACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTG 609
              ||||||||| ||| |||||||||||| ||||| |||||||| |||||||||| ||||| |
Db      74192 ACCTGTAGTTCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGAG 74251

Qy        610 GAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTC 666
              || ||||||||  | | |||||    | |||||||||||||||||  ||||||| ||||||
Db      74252 GAAGTTGCAGTGAGCTGAGATCACGCCACTGCACTCCAGCCTGGTGACAGAGCGAGACTC 74311

Qy        667 CATAAAGACA 676
              || | | |||
Db      74312 CACACACACA 74321


RESULT 5
AAS43104/c
```

```
ID   AAS43104 standard; DNA; 325791 BP.
XX
AC   AAS43104;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human Oestrogen receptor beta gene.
XX
KW   Human; Oestrogen receptor beta; ERbeta; ds; SNP; chromosome 6q.25.1;
KW   single nucleotide polymorphism; cardiovascular disease;
KW   autoimmune disease; systemic lupus erythematosus; arthritis; rheumatism;
KW   osteoarthritis; osteoporosis; breast cancer; endometrial cancer.
XX
OS   Homo sapiens.
XX
PN   WO200162793-A2.
XX
PD   30-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005360.
XX
PR   22-FEB-2000; 2000US-0183755P.
PR   24-JAN-2001; 2001US-00768185.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Kalush F,  Cassel MJ,  Hwang SS,  Winn-Deen ES;
XX
DR   WPI; 2001-582041/65.
DR   P-PSDB; AAU27322.
XX
PT   Estrogen receptor gene and protein polymorphisms useful for diagnosis of
PT   individuals at risk of developing bone disorders.
XX
PS   Example 2; Fig 1; 245pp; English.
XX
CC   The invention relates to a novel isolated peptide comprising or
CC   consisting of an amino acid sequence selected from an amino acid sequence
CC   of a variant oestrogen receptor protein (e.g. ERbeta), or a fragment of
CC   10 amino acids), antibodies against them, nucleic acids encoding them
CC   (including vectors for transforming cells). The gene for human ERbeta is
CC   located on chromosome 6q.25.1. The variants are encoded by single
CC   nucleotide polymorphisms (SNP). The variant peptides and proteins can be
CC   used in assays to determine the biological activity of the protein, to
CC   raise antibodies, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, to identify
CC   compounds that modulate receptor activity and to screen compounds for the
CC   ability to stimulate or inhibit interaction between the receptor protein
CC   and a target molecule that normally interacts with the receptor protein
CC   e.g. oestrogen. The antibody can be used to isolate the protein, to
CC   assess expression in disease states e.g. cardiovascular disease and
CC   autoimmune disease (e.g. systemic lupus erythematosus, arthritis,
CC   rheumatism and osteoarthritis), osteoporosis, breast cancer and
CC   endometrial cancer. In addition the antibodies can be used in
CC   pharmacogenomic analysis and inhibiting protein function, e.g. blocking
CC   the binding of the oestrogen receptor protein to a binding partner such
CC   as a ligand. The nucleic acids encoding the proteins can be used as
CC   probes, primers, chemical intermediates and in biological assays. The
CC   present sequence is the human ERbeta gene
XX
SQ   Sequence 325791 BP; 94098 A; 68292 C; 67970 G; 95431 T; 0 U; 0 Other;

  Query Match             10.7%;  Score 214.4;  DB 4;  Length 325791;
  Best Local Similarity   73.0%;  Pred. No. 6.7e-35;
  Matches  305;  Conservative    0;  Mismatches  106;  Indels    7;  Gaps    2;

Qy        268 TGTTTGATATGAAATAAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTA 327
              || |||| | || | |    | |||||||||  |       ||| || ||| ||  |
Db     257049 TGATTGAGAAGATAGATATTTATCAACAAGTCTTCAAACAAATTTAAAATTAAAATGAGG 256990

Qy        328 TTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTT 387
              |   | | |  |    ||  |||   | ||| | |   |   || |||  ||||   ||
Db     256989 ATACATTCTATGAGAGCATTTAATAGCAGAAGGTGACTTAGGAAAGTTAGAAAAGGCTTT 256930

Qy        388 TGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGT 447
              |||| |||||| ||| ||| |||||||||| |||||||||||  |||
Db     256929 ----CCTGGCCATGCGCAGTGGCTTATGCTTGTAATCCCAGCACTTTGGGAGGCTGAGGC 256874

Qy        448 GGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGT 507
              ||| |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db     256873 GGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCTGT 256814

Qy        508 CTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTAC 567
              |||||| |||||  |||||||||||  ||| |||||||| |||||||| ||||| ||||
Db     256813 CTCTACCAAAAATACAAAAATTAGCTGGGTATGGTGGCAGGCACCTGTAATCCCAGCTAC 256754

Qy        568 TCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGA 627
              |||||  ||||| ||||||||||||| |||  |||||| |||||||||||||| | ||
Db     256753 CTGGGAGGCTGAGGCAGGAGAATTGCTTAAACTCAGGAGGCGGAGGTTGCAGTAAGCTGA 256694

Qy        628 GAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||   | | |||||||||||||||| ||| ||||||||||  ||    | || ||||
Db     256693 GATCACGCCCCTGCACTCCAGCCTGGCGACACAGCAAGACTCTATCTCAAAAAAAAAA 256636


RESULT 6
ACC57389
ID   ACC57389 standard; cDNA; 2013 BP.
XX
AC   ACC57389;
XX
DT   27-JUN-2003  (first entry)
XX
DE   Human macroprotein 1503-8.8 encoding cDNA.
XX
KW   Human; macroprotein; 1503-8.8; dementia; diabetes; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             911. .1153
FT                   /*tag=  a
FT                   /product= "1503-8.8"
XX
PN   CN1359929-A.
XX
PD   24-JUL-2002.
XX
PF   20-DEC-2000; 2000CN-00135135.
XX
PR   20-DEC-2000; 2000CN-00135135.
XX
PA   (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2002-733617/80.
DR   P-PSDB; ABP60188.
XX
PT   Polypeptide-human macroprotein 1503-8.8 and polynucleotide for coding it.
XX
PS   Claim 6; p25-26 (disclosure); 33pp; Chinese.
XX
CC   The invention relates to a novel human macroprotein designated 1503-8.8.
CC   Also disclosed are the polynucleotide encoding the polypeptide, and the
CC   process for preparing the polypeptide using DNA recombination. The
CC   application of the polypeptide is in treating diseases such as dementia
CC   and diabetes. The current sequence represents the human macroprotein 1503
CC   -8.8 encoding cDNA
XX
SQ   Sequence 2013 BP; 622 A; 373 C; 391 G; 627 T; 0 U; 0 Other;

  Query Match             10.7%;  Score 213.6;  DB 6;  Length 2013;
  Best Local Similarity   85.6%;  Pred. No. 4.8e-35;
  Matches  250;  Conservative    0;  Mismatches   39;  Indels    3;  Gaps    1;
```

```
Qy        392 ATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGC 451
              || |||||| | |||| |||| ||||||||||||||| || ||||||||||  ||| ||||
Db        923 ATTGGCCAGACACAGTGGCTCATGCCTGTAATCCCAGCATTTTGGGAGGCTGAGGCGGGC 982

Qy        452 GGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCT 511
              ||||||| |||||||||||||| |||||||||||||||||||| || ||||||||||||
Db        983 GGATCACTTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGCGAGACCCTGTCTCT 1042

Qy        512 ACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAG 571
              ||||||||  |||||||||||||||| |||||||||||||||||| ||||| |||||| |
Db       1043 ACTAAAAATACAAAAATTAGCCAGGTATGGTGGCATGCACCTGTAATCCCAGCTACTCGG 1102

Qy        572 GAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA-- 629
              ||||  |||| ||||||||  ||||||||  ||||| |||||||||||| | || |||
Db       1103 GAGGCAGAGGCAGGAGAATCACTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCAAGATT 1162

Qy        630 -TGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAA 680
               || |||||||||||||||||| ||||||||||||||||    | || ||
Db       1163 GTGCCACTGCACTCCAGCCTGGCGACAGAGCAAGACTCCATCTCAAAAATAA 1214
```

RESULT 7
ABZ70982/c
ID ABZ70982 standard; DNA; 11173 BP.
XX
AC ABZ70982;
XX
DT 28-APR-2003 (first entry)
XX
DE Human HKR1 related DNA sequence SEQ ID NO:10.
XX
KW Human; HKR1; cytostatic; HKR1 inhibitor; hyperproliferative disorder;
KW cancer; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003004513-A1.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US021090.
XX
PR 03-JUL-2001; 2001US-00898556.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett FC, Freier SM;
XX
DR WPI; 2003-210336/20.
XX
PT New compounds, particularly antisense oligonucleotides targeted to a
PT nucleic acid encoding HKR1, useful for treating a disease/condition
PT associated with HKR1, such as hyperproliferative disorder, e.g. lung,
PT brain or breast cancer.
XX
PS Example 15; Page 82-88; 105pp; English.
XX
CC The present invention describes a compound 8-50 nucleobases in length
CC targeted to, and which specifically hybridises with a nucleic acid
CC molecule encoding HKR1, and inhibits the expression of HKR1. Also
CC described: (1) a compound 8-50 nucleobases in length that specifically
CC hybridises with at least an 8-nucleobase portion of an active site on a
CC nucleic acid molecule encoding HKR1; (2) a composition comprising the
CC compound and a carrier or diluent; (3) a method for inhibiting the
CC expression of HKR1 in cells or tissues by contacting the cells or tissues
CC with the compound so that expression of HKR1 is inhibited; and (4) a
CC method of treating an animal having a disease or condition associated
CC with HKR1 by administering to the animal a therapeutic or prophylactic
CC amount of the compound so that expression of HKR1 is inhibited. HKR1
CC antisense oligonucleotides have cytostatic activities and can be used as
CC HKR1 inhibitors. The compound, composition and methods are useful for
CC treating a disease or condition associated with HKR1, such as a
CC hyperproliferative disorder, e.g. lung, brain or breast cancer, by
CC inhibiting the expression of HKR1. They are also useful in research and
CC diagnostics for modulating the expression of HKR1. The present sequence
CC represents a human HKR1 related DNA sequence, which is used in an example
CC from the present invention
XX
SQ Sequence 11173 BP; 2631 A; 2378 C; 2300 G; 3864 T; 0 U; 0 Other;

  Query Match   10.7%; Score 213.2; DB 9; Length 11173;
  Best Local Similarity 81.2%; Pred. No. 7.4e-35;
  Matches 273; Conservative   0; Mismatches   58; Indels   5; Gaps   2;

```
Qy        350 ATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAG 409
              ||||||  ||  || | |   ||||  ||||  | |  ||||| |||||||||   || |
Db       5873 ATTTTACTAAAAAAGATATGCAAATGGTTAATAAACAATGAAAAGGGCCAGGCATGGTGG 5814

Qy        410 CTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGG 469
              ||| |||||||||||||| |||||||||||||   ||| ||| |||||||   ||||||||
Db       5813 CTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCAC--AAGGTCAGG 5756

Qy        470 AGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATT 529
              || || |||||| |||||||||||||||||||||| ||||||||||||    |||||||
Db       5755 AGATTGAGACCATCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAATACAAAAAATT 5696

Qy        530 AGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAA 589
              ||||||||||||||  |||||||||||||| |||||||||||| ||||| ||| |||
Db       5695 AGCCAGGTGTGGTGGTGTGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGGGAA 5636

Qy        590 TTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGC 646
              |||||||||| ||||||  |||||||||||| | |||| ||     ||||| | ||||||
Db       5635 TTGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCGATATAACACCACTGTATTCCAGC 5576

Qy        647 CTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||||||||||||| ||||||  | | || ||||
Db       5575 CTGGCAACAGAGCAAGGCTCCATCTAAAAAAAAAAA 5540
```

RESULT 8
ACN44390
ID ACN44390 standard; DNA; 59725 BP.
XX
AC ACN44390;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG18549.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 814; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of

```
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 59725 BP; 14028 A; 15992 C; 16067 G; 13499 T; 0 U; 139 Other;

  Query Match             10.6%;  Score 212.2;  DB 11;  Length 59725;
  Best Local Similarity   81.0%;  Pred. No. 1.5e-34;
  Matches  260;  Conservative    0;  Mismatches   58;  Indels    3;  Gaps    1;

Qy        364 TACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAAT 423
Db      31622 TACACAAAAATGTGTGGAAAAATGTGAGAGCGGCCGGGTGCAGTGGCTCACGCCTATAAT 31681

Qy        424 CCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGC 483
Db      31682 CCTAGCACTTTGGGAGGCCAAGGCGGGTGGATCACCTGAGGTCAGGAGTTTGAGCCCATC 31741

Qy        484 CTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTG 543
Db      31742 CTGGCCAACATGGCGAAACCCCATCTTTACTAAAAATACAAAAATTAGCTGGGCATGGTG 31801

Qy        544 GCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAG 603
Db      31802 GCATGCACCTGTAGTCCCAGCTACTTGGGAGGCTAAGGCAGGAGAATCACTTGAACCCAG 31861

Qy        604 GAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCA 660
Db      31862 GAGGTGGAGGTTGCAGTGAGCCGAGATCGCGCCATTGCATTCCAGCCTGGCAACAGAGCA 31921

Qy        661 AGACTCCATAAAGACAACAAA 681
Db      31922 AGACTCCATCTCAAAAAATAA 31942


RESULT 9
ADC87336/c
ID   ADC87336 standard; DNA; 108316 BP.
XX
AC   ADC87336;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1789.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87337.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1789; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 108316 BP; 27970 A; 24159 C; 25607 G; 30580 T; 0 U; 0 Other;

  Query Match             10.6%;  Score 211.2;  DB 10;  Length 108316;
  Best Local Similarity   67.9%;  Pred. No. 2.7e-34;
  Matches  342;  Conservative    0;  Mismatches  153;  Indels    9;  Gaps    3;

Qy        283 AAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATT 342
Db      42571 AAAAATGTTACCATGTAATGAGACATTATTGTTATTTTCAAGTGAATTAATACATAAATA 42512

Qy        343 TAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGC 402
Db      42511 TTTTACATACTTTTCCATTTTAATTTGTAATACAGTCAAAAATCAATAAACAGGCCAGGC 42452

Qy        403 GCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGA 462
Db      42451 GAGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCGGATCACCTGA 42392

Qy        463 GGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGC 522
Db      42391 GGTCAGGAGTTCGAGACCAGCCTGGACAACATGGTGAAACCCTGTCTCTACTAAAAATAC 42332

Qy        523 AAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGG 582
Db      42331 AAAAATTAGCTGGGCGTGGTGGCGCATGCCTGTAATCCCAGCTACTTGGGAGGCTAAGGC 42272

Qy        583 AGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCA 639
Db      42271 AGGAGAATCGCTTGAGCCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCACACCACTGCA 42212

Qy        640 CTCCAGCCT-GGCAACAGAGCAAGACTCCAT-----AAAGACAACAAAAGCTTTGAAATT 693
Db      42211 CTCCAGCCTGGGCAACAGAGCAAGACTCCATCTCAAAAAAATATGATAAGGTCCAAAATA 42152

Qy        694 GTGTAAATGAGTTGTACCTATCTTCATTTAAGAAATTCATCTTTGTTCATTTATTTTTAC 753
Db      42151 TTTTCTAGAAGATTTGACAATCTAATATTTCTATGTGAGTATCAAGATTTATACTACTCA 42092

Qy        754 TTGACATGAGAGCTTCCAGCAATT 777
Db      42091 TGGCTTTGAAATTTTCAAGAATTT 42068


RESULT 10
AAS41753/c
ID   AAS41753 standard; DNA; 11950 BP.
XX
AC   AAS41753;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Genomic sequence #69 encoding novel human enzyme polypeptide.
XX
KW   Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW   ligase; hyperproliferative disorder; immunodeficiency disorder;
KW   autoimmune disorder; neurological disorder; metabolic disorder;
KW   inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW   blood-related disorder; infectious disorder; gene therapy; cytostatic;
```

KW anti arthritic; nephrotropic; anticoagulant; ds.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001239.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465566/50.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT treating neural, immune system, muscular, reproductive, pulmonary,
PT cardiovascular, renal, proliferative disorders and cancerous diseases.
XX
PS Disclosure; SEQ ID NO 1879; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC infectious disorders (e.g. Influenza). The polynucleotides of the
CC invention can also be used in gene therapy. AAS41685-AAS42192 represent
CC DNA sequences encoding for the novel human enzyme polypeptides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11950 BP; 3159 A; 2617 C; 2790 G; 3384 T; 0 U; 0 Other;

Query Match 10.5%; Score 210.8; DB 4; Length 11950;
Best Local Similarity 59.6%; Pred. No. 2.4e-34;
Matches 409; Conservative 0; Mismatches 272; Indels 5; Gaps 3;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| ||| ||||||||||||||||||||||| || ||||||||  | ||||
Db       9406 TGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACATGCATGAGCCACTGCACCTG 9347

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              |  ||| | |  || ||   || | ||   |        | ||     | |  || ||
Db       9346 GCAAAATTTTTAAATAATAAAAAAACAC-CTAGAATAAAAAGATAAAAATCATCCTTCAC 9288

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | ||    |  ||          |||  | | ||       |  |   |
Db       9287 ACTCTGCTTGGCTCTCCCTAAAACTCCCTGGAAGTAACCACTACTGCCCAGTGGCTGGTT 9228

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
                |  |||     | |||   |||  ||  |   ||| |  ||  || ||
Db       9227 ATCCCTCCAGACTGTTCTCTATGTATAAAAAATCATATAGTTTTGTATTCCAGTTTTTC 9168

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
               ||      | |    | |  | | | |  | |      |||  || | |
Db       9167 CTTACAACATTATATCACTAACATTTTTATATTAATTACATATTATATTAAAATATGATA 9108

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |      |  |   ||||  |  ||||||   || | || | | | | | |  |
Db       9107 TTTAATGACTGTATATTTTATTACATATTTAAGGCAGCAATGTTAAACACTCAGACTCTA 9048

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              | ||          | || | |  |        | ||||||||||| ||||  |||||
Db       9047 GAGTAGCAACTGTCCAATAGAAATATGATGGGAAGACCAGGCGCAGTGGCTCACGCCTGC 8988

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||  |||||||||||||| ||||||||||||| ||||||
Db       8987 AATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACTTGAGGTCAGGAGTTCAAGACC 8928

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||| ||||||||  |||| |  |||||||| |||  |||||| ||||  || ||
Db       8927 AGCCTGGTCAACATGGCAAAACTCCATCTCTACTAGAAATACAAAAACTAGCTGGGCGTC 8868

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              |||||||||||||||| ||||| |||||| ||||| ||||| ||||| ||  ||||||||
Db       8867 GTGGCATGCACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGATCCCTTGAACC 8808

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAG 656
              |||||||||||||||||| | || ||    ||| || ||||||||||||| |||||||
Db       8807 CAGGAGGTGGAGGTTGCAGTGAGCCAAGATCATGCCATTGCACTCCAGCCTGGGCAACAG 8748

Qy        657 AGCAAGACTCCATAAAGACAACAAAA 682
              ||  |||||   |    | || ||||
Db       8747 AGTGAGACTATGTCTCCAAAAAAAAA 8722
```

RESULT 11
ABA06813/c
ID ABA06813 standard; DNA; 11950 BP.
XX
AC ABA06813;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human genomic DNA SEQ ID NO: 899.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ds.
XX
OS Homo sapiens.
XX
PN WO200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001349.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476161/51.
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition.
XX
PS Disclosure; SEQ ID NO 899; 859pp + Sequence Listing; English.
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a genomic DNA of the invention
XX
SQ Sequence 11950 BP; 3159 A; 2617 C; 2790 G; 3384 T; 0 U; 0 Other;

Query Match 10.5%; Score 210.8; DB 4; Length 11950;
Best Local Similarity 59.6%; Pred. No. 2.4e-34;
Matches 409; Conservative 0; Mismatches 272; Indels 5; Gaps 3;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| ||| ||||||||||||||||||||||| || |||||||| | ||||
Db       9406 TGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACATGCATGAGCCACTGCACCTG 9347

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
                |  ||| | |  || ||   || | ||   |         | ||      | |  || ||
Db       9346 GCAAAATTTTTAAATAATAAAAAAACAC-CTAGAATAAAAAGATAAAAATCATCCTTCAC 9288

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              |  ||      |   ||           |||  | | ||          | |   |
Db       9287 ACTCTGCTTGGCTCTCCCTAAAACTCCCTGGAAGTAACCACTACTGCCCAGTGGCTGGTT 9228

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
                |   |||     | |||   |||  ||  |   ||| |  ||  || ||
Db       9227 ATCCCTCCAGACTGTTCTCTATGTATAAAAAATCATATAGTTTTGTATTCCAGTTTTTTC 9168

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
                ||      | |    | |   |  | | | |  |  |      |||  || | |
Db       9167 CTTACAACATTATATCACTAACATTTTTATATTAATTACATATTATATTAAAATATGATA 9108

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
```

```
Db       9107 TTTAATGACTGTATATTTTATTACATATTTAAGGCAGCAATGTTAAACACTCAGACTCTA 9048

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420

Db       9047 GAGTAGCAACTGTCCAATAGAAATATGATGGGAAGACCAGGCGCAGTGGCTCACGCCTGC 8988

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480

Db       8987 AATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACTTGAGGTCAGGAGTTCAAGACC 8928

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540

Db       8927 AGCCTGGTCAACATGGCAAAACTCCATCTCTACTAGAAATACAAAAACTAGCTGGGCGTC 8868

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600

Db       8867 GTGGCATGCACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGATCCCTTGAACC 8808

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAG 656

Db       8807 CAGGAGGTGGAGGTTGCAGTGAGCCAAGATCATGCCATTGCACTCCAGCCTGGGCAACAG 8748

Qy        657 AGCAAGACTCCATAAAGACAACAAAA 682

Db       8747 AGTGAGACTATGTCTCCAAAAAAAAA 8722


RESULT 12
AAS29743/c
ID   AAS29743 standard; DNA; 11950 BP.
XX
AC   AAS29743;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human endocrine polypeptide encoding genomic DNA SEQ ID No 469.
XX
KW   Endocrine protein; human; mouse; rabbit; goat; horse; food additive; cat;
KW   dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; food preservative;
KW   tissue regeneration; anti-infertility.
XX
OS   Homo sapiens.
XX
PN   WO200155364-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001308.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
```

PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451936/48.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders of the endocrine system such as reproductive disorders,
PT endocrine cancers and also for testing and detection e.g. diagnosis.
XX
PS Claim 1; SEQ ID NO 469; 604pp; English.
XX
CC Sequences AAS29737-AAS29769 represent genomic DNA molecules, which encode
CC the endocrine polypeptides of the invention. Endocrine polypeptides and
CC their associated polynucleotides of the invention are useful in the
CC diagnosis, treatment and prevention of various types of disorders in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by determining the presence or
CC absence of a mutation in an endocrine polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma. The polypeptides can also be used
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, to regenerate tissues and in chemotaxis.
CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

Query Match 10.5%; Score 210.8; DB 5; Length 11950;
Best Local Similarity 59.6%; Pred. No. 2.4e-34;
Matches 409; Conservative 0; Mismatches 272; Indels 5; Gaps 3;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||  |||  ||||||||||||||||||||||||| || ||||||||  | ||||
Db       9406 TGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACATGCATGAGCCACTGCACCTG 9347

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              |   ||| | |  || ||   || | ||   |        | ||      |  |  || ||
Db       9346 GCAAAATTTTTAAATAATAAAAAAACAC-CTAGAATAAAAAGATAAAAATCATCCTTCAC 9288

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              |  ||      |   ||           |||  | | ||        |  |   |
Db       9287 ACTCTGCTTGGCTCTCCCTAAAACTCCCTGGAAGTAACCACTACTGCCCAGTGGCTGGTT 9228

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |   |||     | |||   |||  ||  |   ||| |  ||  || ||
Db       9227 ATCCCTCCAGACTGTTCTCTATGTATAAAAAATCATATAGTTTTGTATTCCAGTTTTTTC 9168

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              ||      | |    | |   |  | | | |  | |      |||  || | |
Db       9167 CTTACAACATTATATCACTAACATTTTTATATTAATTACATATTATATTAAAATATGATA 9108

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |         |  |   ||||  |  ||||||   || | || | | | | | |
Db       9107 TTTAATGACTGTATATTTTATTACATATTTAAGGCAGCAATGTTAAACACTCAGACTCTA 9048

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              | ||          | || | |  |        | ||||||||||| ||||  |||||
Db       9047 GAGTAGCAACTGTCCAATAGAAATATGATGGGAAGACCAGGCGCAGTGGCTCACGCCTGC 8988

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||  |||||||||||||||  |||||||||||||||  ||||||
Db       8987 AATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACTTGAGGTCAGGAGTTCAAGACC 8928

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||| ||||||||  ||||  |  |||||||| |||  |||||  ||||  || ||
Db       8927 AGCCTGGTCAACATGGCAAAACTCCATCTCTACTAGAAATACAAAAACTAGCTGGGCGTC 8868

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              |||||||||||||||  |||||  ||||||  |||||  |||||  |||||  ||  ||||||||
Db       8867 GTGGCATGCACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGATCCCTTGAACC 8808

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAG 656
              |||||||||||||||||| | || ||   ||| || ||||||||||||| ||||||||
Db       8807 CAGGAGGTGGAGGTTGCAGTGAGCCAAGATCATGCCATTGCACTCCAGCCTGGGCAACAG 8748

Qy        657 AGCAAGACTCCATAAAGACAACAAAA 682
              ||  |||||    |    | || ||||
Db       8747 AGTGAGACTATGTCTCCAAAAAAAAA 8722
```

RESULT 13
ABV84150/c
ID ABV84150 standard; DNA; 11950 BP.
XX
AC ABV84150;
XX
DT 09-DEC-2002 (first entry)
XX

DE Human polynucleotide SEQ ID NO 899.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN US2002090672-A1.
XX
PD 11-JUL-2002.
XX
PF 17-JAN-2001; 2001US-00764853.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2002-681727/73.
XX
PT Novel polypeptide useful for diagnosis, prognosis, prevention, and
PT treatment of immune, hyperproliferative, renal, respiratory,
PT cardiovascular, reproductive, endocrine, gastrointestinal and
PT neurological disorders.
XX
PS Disclosure; SEQ ID NO 899; 369pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABV83682-ABV84101) and proteins
CC (ABP66710-ABP67129) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11950 BP; 3159 A; 2617 C; 2790 G; 3384 T; 0 U; 0 Other;

Query Match 10.5%; Score 210.8; DB 6; Length 11950;
Best Local Similarity 59.6%; Pred. No. 2.4e-34;
Matches 409; Conservative 0; Mismatches 272; Indels 5; Gaps 3;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||||||  |||  |||||||||||||||||||||||| || ||||||||  | ||||
Db       9406 TGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACATGCATGAGCCACTGCACCTG 9347

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              |   ||| | |  || ||   || | ||   |        | ||      | |  || ||
Db       9346 GCAAAATTTTTAAATAATAAAAAAACAC-CTAGAATAAAAAGATAAAAATCATCCTTCAC 9288

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | ||      |  ||           |||  | | ||        | |  |
Db       9287 ACTCTGCTTGGCTCTCCCTAAAACTCCCTGGAAGTAACCACTACTGCCCAGTGGCTGGTT 9228

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |  |||     | |||  |||  ||  |  ||| |  ||  || ||
Db       9227 ATCCCTCCAGACTGTTCTCTATGTATAAAAAATCATATAGTTTTGTATTCCAGTTTTTTC 9168

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              ||      | |   | |  |  | | | |  |  |      |||  || | |
Db       9167 CTTACAACATTATATCACTAACATTTTTATATTAATTACATATTATATTAAAATATGATA 9108

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |         |  |  ||||  |  ||||||   || | || | | | | | |  |
Db       9107 TTTAATGACTGTATATTTTATTACATATTTAAGGCAGCAATGTTAAACACTCAGACTCTA 9048

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              | ||         | || | |  |       | |||||||||| ||||  |||||
Db       9047 GAGTAGCAACTGTCCAATAGAAATATGATGGGAAGACCAGGCGCAGTGGCTCACGCCTGC 8988

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||  |||||||||||||| |||||||||||||| ||||||
Db       8987 AATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACTTGAGGTCAGGAGTTCAAGACC 8928

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||| ||||||||  ||||  |  ||||||||| |||  |||||| ||||  || ||
Db       8927 AGCCTGGTCAACATGGCAAAACTCCATCTCTACTAGAAATACAAAAACTAGCTGGGCGTC 8868

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
```

```
              ||||||||||||||||| ||||| |||||| ||||| ||||| ||||| ||  ||||||||
Db       8867 GTGGCATGCACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGATCCCTTGAACC 8808

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAG 656
              ||||||||||||||||||||| | || ||   ||| || |||||||||||||| ||||||||
Db       8807 CAGGAGGTGGAGGTTGCAGTGAGCCAAGATCATGCCATTGCACTCCAGCCTGGGCAACAG 8748

Qy        657 AGCAAGACTCCATAAAGACAACAAAA 682
              ||  |||||    |     | || ||||
Db       8747 AGTGAGACTATGTCTCCAAAAAAAAA 8722


RESULT 14
ADE43582/c
ID   ADE43582 standard; DNA; 128034 BP.
XX
AC   ADE43582;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Polymorphic human IDE genomic sequence, SEQ ID 187.
XX
KW   Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;
KW   Alzheimer's disease; neuroprotective; nootropic; gene therapy;
KW   Chromosome 10; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    2456
FT                   /*tag=  a
FT                   /note= "There is a variation at this position"
FT   misc_feature    3279
FT                   /*tag=  b
FT                   /note= "There is a variation at this position"
FT   misc_feature    3407
FT                   /*tag=  c
FT                   /note= "There is a variation at this position"
FT   misc_feature    17095
FT                   /*tag=  d
FT                   /note= "There is a variation at this position"
FT   misc_feature    17242
FT                   /*tag=  e
FT                   /note= "There is a variation at this position"
FT   misc_feature    33590
FT                   /*tag=  f
FT                   /note= "There is a variation at this position"
FT   misc_feature    38903
FT                   /*tag=  g
FT                   /note= "There is a variation at this position"
FT   misc_feature    42943
FT                   /*tag=  h
FT                   /note= "There is a variation at this position"
FT   misc_feature    43391
FT                   /*tag=  i
FT                   /note= "There is a variation at this position"
FT   misc_feature    45017
FT                   /*tag=  j
FT                   /note= "There is a variation at this position"
FT   misc_feature    62498
FT                   /*tag=  k
FT                   /note= "There is a variation at this position"
FT   misc_feature    68906
FT                   /*tag=  l
FT                   /note= "There is a variation at this position"
FT   misc_feature    68973
FT                   /*tag=  m
FT                   /note= "There is a variation at this position"
FT   misc_feature    69586
FT                   /*tag=  n
FT                   /note= "There is a variation at this position"
FT   misc_feature    73772
FT                   /*tag=  o
FT                   /note= "There is a variation at this position"
FT   misc_feature    74084
FT                   /*tag=  p
FT                   /note= "There is a variation at this position"
FT   misc_feature    83024
FT                   /*tag=  q
FT                   /note= "There is a variation at this position"
FT   misc_feature    83104
FT                   /*tag=  r
FT                   /note= "There is a variation at this position"
FT   misc_feature    89301
FT                   /*tag=  s
FT                   /note= "There is a variation at this position"
FT   misc_feature    105060
FT                   /*tag=  t
FT                   /note= "There is a variation at this position"
FT   misc_feature    107395
FT                   /*tag=  u
FT                   /note= "There is a variation at this position"
FT   misc_feature    108489
FT                   /*tag=  v
FT                   /note= "There is a variation at this position"
FT   misc_feature    111914
FT                   /*tag=  w
FT                   /note= "There is a variation at this position"
FT   misc_feature    112114
FT                   /*tag=  x
FT                   /note= "There is a variation at this position"
FT   misc_feature    113142
FT                   /*tag=  y
FT                   /note= "There is a variation at this position"
FT   misc_feature    113591
FT                   /*tag=  z
FT                   /note= "There is a variation at this position"
FT   misc_feature    114683
FT                   /*tag=  aa
FT                   /note= "There is a variation at this position"
FT   misc_feature    116662
FT                   /*tag=  ab
FT                   /note= "There is a variation at this position"
FT   misc_feature    117803
FT                   /*tag=  ac
FT                   /note= "There is a variation at this position"
FT   misc_feature    124565
FT                   /*tag=  ad
FT                   /note= "There is a variation at this position"
XX
PN   WO2003054143-A2.
XX
PD   03-JUL-2003.
XX
PF   25-OCT-2002; 2002WO-US034679.
XX
PR   25-OCT-2001; 2001US-0339525P.
PR   08-NOV-2001; 2001US-0336929P.
PR   08-NOV-2001; 2001US-0338010P.
PR   09-NOV-2001; 2001US-0338363P.
PR   04-DEC-2001; 2001US-0337052P.
PR   28-MAR-2002; 2002US-0368919P.
XX
PA   (NEUR-) NEUROGENETICS INC.
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Becker KD,  Velicelebi G,  Elliott KJ,  Wang X,  Tanzi RE,  Bertram L;
PI   Saunders AJ,  Mullin KM,  Sampson AJ,  Blacker DL;
XX
DR   WPI; 2003-559131/52.
XX
PT   Determining a predisposition for or the occurrence of neurodegenerative
PT   disease, e.g. Alzheimer's disease by detecting in a target nucleic acid
PT   the presence or absence of an allelic variant of one or more polymorphic
PT   regions.
XX
```

PS Claim 9; Page 584-page 618; 848pp; English.
XX
CC The present invention relates to a method (M1) for determining a
CC predisposition for or the occurrence of neurodegenerative disease in a
CC subject. The method comprises detecting in a target nucleic acid obtained
CC from the subject the presence or absence of an allelic variant of one or
CC more polymorphic regions of one or more genes selected from uPA
CC (Urokinase plasminogen activator), SNCG (gamma-synuclein), IDE (insulin-
CC degrading enzyme), KNSL1 (Kinesin-like protein 1), LIPA (lysosomal acid
CC lypase), and TNFRSF6 (Tumour Necrosis Factor Receptor-SF6), where the
CC presence of at least one of the allelic variant of one or more
CC polymorphic regions is indicative of a predisposition for or the
CC occurrence of neurodegenerative disease. The genes are all located on
CC chromosome 10. M1 is useful for determining a predisposition for or the
CC occurrence of, and for treating neurodegenerative disease, particularly
CC Alzheimer's disease.
XX
SQ Sequence 128034 BP; 34726 A; 25977 C; 26400 G; 40799 T; 0 U; 132 Other;

Query Match 10.5%; Score 210.6; DB 10; Length 128034;
Best Local Similarity 60.6%; Pred. No. 3.7e-34;
Matches 433; Conservative 0; Mismatches 274; Indels 8; Gaps 5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| ||||  ||||| ||||||||||||||||||||||||||||||||||||||
Db      81438 TGATCCACCTGCCTCAGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 81379

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | | || || ||| |||    |||| | ||     |    | ||  | |||  | || |
Db      81378 GCCTAAAGTGTTTCTTAAAGTAAGAAAAAAGTCAAG--AAAGATTAGGCTTAAAGTTTAG 81321

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | | ||   |  | |   |      | | |||  |     | |    | |  |  ||
Db      81320 AAAATTCAGTCAACTTTTACTGAGTTCTTACCAGGTCAAATTCTGGGAAAGACAAAATAT 81261

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | ||  |  ||        ||||  | |   |||  ||||    |  |||   | ||
Db      81260 AGT-TCCTATTTGGAAAAACTCATAATTTAATAGGACACATGACCTATAGCAAAATGTAA 81202

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              | |    ||   | ||    ||       |     |          | |     | |
Db      81201 TCAGTTAATCTCATGTTAACAGATGGCAGGACTACTCCTCTCTTATTTACCTGGTCTTTG 81142

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTA-CAACTATTTAATAAAATTTTAAAAA 359
              |      || |   || ||  || |    | |||  |   || |||    |  ||
Db      81141 AAATAAACATGGCTCTTTAAACTGTAAAGGTGGACCAAACACAGAAAAAAGGAGTTGAAG 81082

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              ||    | | |    | ||   | |          ||||| |  || |||| ||||||
Db      81081 ATAGGTGAATGAGCCAAGTATGAAACCACCTCTCTAGCCAGGTGTGGTGGCTCATGCCTG 81022

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              ||||| ||||| |||||||||||  |||  || |||||||  ||||||||||||| |||||
Db      81021 TAATCTCAACAATTTGGGAGGCTGAGGCCGGTGGATCACTTGAGGTCAGGAGTTCAAGAC 80962

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              |||||||||||| |||||||||||| |||||||||||  |||||||||||||||||||
Db      80961 CAGCCTGGCCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCCAGGTGT 80902

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||| |   ||||||||| | ||| | |||| ||||| ||||| ||||||||   ||||||
Db      80901 GGTAGTGGGCACCTGTAATGCCAGCCACTCGGGAGGCTGAGGCAGGAGAATCCTTTGAAC 80842

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAACA 655
              | ||||||  |||| |||||| | ||||||   || || ||||||||||||| |||||||
Db      80841 CCAGGAGGCCGAGGCTGCAGTGAGCCGAGATTGTGCCATTGCACTCCAGCCTGGGCAACA 80782

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTAC 710
              |||||||||||  |    | ||  | |    | |||  |   |||| |  | | |
Db      80781 GAGCAAGACTCTGTCTCAAAAAATATAAAAATAAAAAAGGAGAAATCACATCTCC 80727
```

RESULT 15

ADE43581/c
ID ADE43581 standard; DNA; 128034 BP.
XX
AC ADE43581;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human IDE genomic sequence, SEQ ID 186.
XX
KW Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;
KW Alzheimer's disease; neuroprotective; nootropic; gene therapy;
KW Chromosome 10; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003054143-A2.
XX
PD 03-JUL-2003.
XX
PF 25-OCT-2002; 2002WO-US034679.
XX
PR 25-OCT-2001; 2001US-0339525P.
PR 08-NOV-2001; 2001US-0336929P.
PR 08-NOV-2001; 2001US-0338010P.
PR 09-NOV-2001; 2001US-0338363P.
PR 04-DEC-2001; 2001US-0337052P.
PR 28-MAR-2002; 2002US-0368919P.
XX
PA (NEUR-) NEUROGENETICS INC.
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Becker KD, Velicelebi G, Elliott KJ, Wang X, Tanzi RE, Bertram L;
PI Saunders AJ, Mullin KM, Sampson AJ, Blacker DL;
XX
DR WPI; 2003-559131/52.
XX
PT Determining a predisposition for or the occurrence of neurodegenerative
PT disease, e.g. Alzheimer's disease by detecting in a target nucleic acid
PT the presence or absence of an allelic variant of one or more polymorphic
PT regions.
XX
PS Claim 22; Page 549-584; 848pp; English.
XX
CC The present invention relates to a method (M1) for determining a
CC predisposition for or the occurrence of neurodegenerative disease in a
CC subject. The method comprises detecting in a target nucleic acid obtained
CC from the subject the presence or absence of an allelic variant of one or
CC more polymorphic regions of one or more genes selected from uPA
CC (Urokinase plasminogen activator), SNCG (gamma-synuclein), IDE (insulin-
CC degrading enzyme), KNSL1 (Kinesin-like protein 1), LIPA (lysosomal acid
CC lypase), and TNFRSF6 (Tumour Necrosis Factor Receptor-SF6), where the
CC presence of at least one of the allelic variant of one or more
CC polymorphic regions is indicative of a predisposition for or the
CC occurrence of neurodegenerative disease. The genes are all located on
CC chromosome 10. M1 is useful for determining a predisposition for or the
CC occurrence of, and for treating neurodegenerative disease, particularly
CC Alzheimer's disease.
XX
SQ Sequence 128034 BP; 34731 A; 25985 C; 26409 G; 40808 T; 0 U; 101 Other;

Query Match 10.5%; Score 210.6; DB 10; Length 128034;
Best Local Similarity 60.6%; Pred. No. 3.7e-34;
Matches 433; Conservative 0; Mismatches 274; Indels 8; Gaps 5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| ||||  ||||| ||||||||||||||||||||||||||||||||||||||
Db      81438 TGATCCACCTGCCTCAGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 81379

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | | || || ||| |||    |||| | ||     |    | ||  | |||  | || |
Db      81378 GCCTAAAGTGTTTCTTAAAGTAAGAAAAAAGTCAAG--AAAGATTAGGCTTAAAGTTTAG 81321

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
```

```
Db      81320 AAAATTCAGTCAACTTTTACTGAGTTCTTACCAGGTCAAATTCTGGGAAAGACAAAATAT 81261

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240

Db      81260 AGT-TCCTATTTGGAAAAACTCATAATTTAATAGGACACATGACCTATAGCAAAATGTAA 81202

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300

Db      81201 TCAGTTAATCTCATGTTAACAGATGGCAGGACTACTCCTCTCTTATTTACCTGGTCTTTG 81142

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTA-CAACTATTTAATAAAATTTTAAAAA 359

Db      81141 AAATAAACATGGCTCTTTAAACTGTAAAGGTGGACCAAACACAGAAAAAAGGAGTTGAAG 81082

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419

Db      81081 ATAGGTGAATGAGCCAAGTATGAAACCACCTCTCTAGCCAGGTGTGGTGGCTCATGCCTG 81022

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479

Db      81021 TAATCTCAACAATTTGGGAGGCTGAGGCCGGTGGATCACTTGAGGTCAGGAGTTCAAGAC 80962

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539

Db      80961 CAGCCTGGCCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCCAGGTGT 80902

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599

Db      80901 GGTAGTGGGCACCTGTAATGCCAGCCACTCGGGAGGCTGAGGCAGGAGAATCCTTTGAAC 80842

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAACA 655

Db      80841 CCAGGAGGCCGAGGCTGCAGTGAGCCGAGATTGTGCCATTGCACTCCAGCCTGGGCAACA 80782

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTAC 710

Db      80781 GAGCAAGACTCTGTCTCAAAAAATATAAAAATAAAAAAGGAGAAATCACATCTCC 80727
```

```
RESULT 16
ADH54059/c
ID   ADH54059 standard; DNA; 128034 BP.
XX
AC   ADH54059;
XX
DT   25-MAR-2004  (first entry)
XX
DE   Human IDE gene DNA sequence SeqID186.
XX
KW   human; neurodegenerative disease; urokinase plasminogen activator; uPA;
KW   gamma-synuclein; SNCG; insulin degrading enzyme; IDE;
KW   kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;
KW   tumour necrosis factor receptor SF6; TNFRSF6; Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(2456,G)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3279,C)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3407,T)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17095,G)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17242,T)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(33590,G)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(38903,T)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(42943,C)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(43391,G)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(45017,T)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(62498,C)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(68906,G)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(68973,T)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(69586,C)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(73772,G)
FT                   /*tag=  o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(74084,T)
FT                   /*tag=  p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(83024,T)
FT                   /*tag=  q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(83104,T)
FT                   /*tag=  r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(89301,G)
FT                   /*tag=  s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(105060,G)
FT                   /*tag=  t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(107395,A)
FT                   /*tag=  u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(108489,T)
FT                   /*tag=  v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(111914,G)
FT                   /*tag=  w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(112114,A)
FT                   /*tag=  x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(113142,G)
FT                   /*tag=  y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(113591,G)
FT                   /*tag=  z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(114683,G)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(116662,A)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(117803,T)
FT                   /*tag=  ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(124565,T)
```

```
FT                   /*tag=  ad
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   US2003224380-A1.
XX
PD   04-DEC-2003.
XX
PF   25-OCT-2002; 2002US-00282174.
XX
PR   25-OCT-2001; 2001US-0339525P.
PR   25-OCT-2001; 2001US-0348065P.
PR   02-NOV-2001; 2001US-0336983P.
PR   08-NOV-2001; 2001US-0336929P.
PR   08-NOV-2001; 2001US-0338010P.
PR   09-NOV-2001; 2001US-0338363P.
PR   04-DEC-2001; 2001US-0337052P.
PR   28-MAR-2002; 2002US-0368919P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Becker KD,  Velicelebi G,  Ellliott KJ,  Wang X,  Tanzi RE;
PI   Bertram L,  Saunders AJ,  Mullin KM,  Sampson AJ;
XX
DR   WPI; 2004-060538/06.
XX
PT   Determining a predisposition for or the occurrence of neurodegenerative
PT   disease, particularly Alzheimer's disease, comprises determining the
PT   presence of a polymorphism in the uPA, SNCG, IDE, KNSL1, LIPA or TNFRSF6
PT   gene.
XX
PS   Claim 22; SEQ ID NO 186; 205pp; English.
XX
CC   This invention relates to a novel method of determining a predisposition
CC   for or the occurrence of neurodegenerative disease comprising detecting
CC   in a target nucleic acid obtained from the subject the presence of an
CC   allelic variant of polymorphic regions of human genes selected from
CC   urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin
CC   degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid
CC   lipase (LIPA) and tumour necrosis factor receptor SF6 (TNFRSF6). The
CC   method is useful in determining the presence or predisposition to a
CC   neurodegenerative disease, particularly Alzheimer's disease. The present
CC   sequence is that of the human IDE gene which is related to the invention.
XX
SQ   Sequence 128034 BP; 34731 A; 25985 C; 26409 G; 40808 T; 0 U; 101 Other;

  Query Match             10.5%;  Score 210.6;  DB 12;  Length 128034;
  Best Local Similarity   60.6%;  Pred. No. 3.7e-34;
  Matches  433;  Conservative    0;  Mismatches  274;  Indels    8;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| ||||  ||||| ||||||||||||||||||||||||||||||||||||
Db      81438 TGATCCACCTGCCTCAGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 81379

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | | || || ||| |||    |||| | ||     |   | ||  | |||  | || |
Db      81378 GCCTAAAGTGTTTCTTAAAGTAAGAAAAAAGTCAAG--AAAGATTAGGCTTAAAGTTTAG 81321

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | | ||   | | |    |       | | ||| |       | |    | |  |  ||
Db      81320 AAAATTCAGTCAACTTTTACTGAGTTCTTACCAGGTCAAATTCTGGGAAAGACAAAATAT 81261

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | ||  |  ||      |||| | |  |||  ||||   |  |||   | ||
Db      81260 AGT-TCCTATTTGGAAAAACTCATAATTTAATAGGACACATGACCTATAGCAAAATGTAA 81202

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              | |    ||   | ||    ||       |    |             |  |      | |
Db      81201 TCAGTTAATCTCATGTTAACAGATGGCAGGACTACTCCTCTCTTATTTACCTGGTCTTTG 81142

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTA-CAACTATTTAATAAAATTTTAAAAA 359
              |      || |    || ||  || |       | |||  |   || |||    |  ||
Db      81141 AAATAAACATGGCTCTTTAAACTGTAAAGGTGGACCAAACACAGAAAAAAGGAGTTGAAG 81082

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              ||    | | |    | ||   | |          ||||| |  || |||| ||||||
Db      81081 ATAGGTGAATGAGCCAAGTATGAAACCACCTCTCTAGCCAGGTGTGGTGGCTCATGCCTG 81022

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              ||||| |||| |||||||||| ||| || |||||| |||||||||||||| |||||
Db      81021 TAATCTCAACAATTTGGGAGGCTGAGGCCGGTGGATCACTTGAGGTCAGGAGTTCAAGAC 80962

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              |||||||||| |||||||||| |||||||||||  |||||||||||||||||||
Db      80961 CAGCCTGGCCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCCAGGTGT 80902

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||| |   ||||||||| | ||| | |||| ||||| ||||| ||||||||    ||||||
Db      80901 GGTAGTGGGCACCTGTAATGCCAGCCACTCGGGAGGCTGAGGCAGGAGAATCCTTTGAAC 80842

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAACA 655
              | |||||| |||| |||||| | ||||||   || || ||||||||||||| |||||||
Db      80841 CCAGGAGGCCGAGGCTGCAGTGAGCCGAGATTGTGCCATTGCACTCCAGCCTGGGCAACA 80782

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTAC 710
              ||||||||||  |    | ||  | |    | |||  |   |||| |  | | |
Db      80781 GAGCAAGACTCTGTCTCAAAAAATATAAAAATAAAAAAGGAGAAATCACATCTCC 80727


RESULT 17
ADH54060/c
ID   ADH54060 standard; DNA; 128034 BP.
XX
AC   ADH54060;
XX
DT   25-MAR-2004  (first entry)
XX
DE   Human IDE gene variant DNA sequence SeqID187.
XX
KW   human; neurodegenerative disease; urokinase plasminogen activator; uPA;
KW   gamma-synuclein; SNCG; insulin degrading enzyme; IDE;
KW   kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;
KW   tumour necrosis factor receptor SF6; TNFRSF6; Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
PN   US2003224380-A1.
XX
PD   04-DEC-2003.
XX
PF   25-OCT-2002; 2002US-00282174.
XX
PR   25-OCT-2001; 2001US-0339525P.
PR   25-OCT-2001; 2001US-0348065P.
PR   02-NOV-2001; 2001US-0336983P.
PR   08-NOV-2001; 2001US-0336929P.
PR   08-NOV-2001; 2001US-0338010P.
PR   09-NOV-2001; 2001US-0338363P.
PR   04-DEC-2001; 2001US-0337052P.
PR   28-MAR-2002; 2002US-0368919P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Becker KD,  Velicelebi G,  Ellliott KJ,  Wang X,  Tanzi RE;
PI   Bertram L,  Saunders AJ,  Mullin KM,  Sampson AJ;
XX
DR   WPI; 2004-060538/06.
XX
PT   Determining a predisposition for or the occurrence of neurodegenerative
PT   disease, particularly Alzheimer's disease, comprises determining the
PT   presence of a polymorphism in the uPA, SNCG, IDE, KNSL1, LIPA or TNFRSF6
PT   gene.
XX
PS   Claim 9; SEQ ID NO 187; 205pp; English.
XX
CC   This invention relates to a novel method of determining a predisposition
CC   for or the occurrence of neurodegenerative disease comprising detecting
```

```
CC   in a target nucleic acid obtained from the subject the presence of an
CC   allelic variant of polymorphic regions of human genes selected from
CC   urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin
CC   degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid
CC   lipase (LIPA) and tumour necrosis factor receptor SF6 (TNFRSF6). The
CC   method is useful in determining the presence or predisposition to a
CC   neurodegenerative disease, particularly Alzheimer's disease. The present
CC   sequence is that of the human IDE gene, with polymorphic sites
CC   represented by n, which is related to the invention.
XX
SQ   Sequence 128034 BP; 34726 A; 25977 C; 26400 G; 40799 T; 0 U; 132 Other;

  Query Match            10.5%;  Score 210.6;  DB 12;  Length 128034;
  Best Local Similarity  60.6%;  Pred. No. 3.7e-34;
  Matches  433;  Conservative    0;  Mismatches  274;  Indels    8;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| ||||  ||||| ||||||||||||||||||||||||||||||||||||||
Db      81438 TGATCCACCTGCCTCAGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 81379

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | | || || ||| |||    |||| | ||      |   | ||  | |||  | || |
Db      81378 GCCTAAAGTGTTTCTTAAAGTAAGAAAAAAGTCAAG--AAAGATTAGGCTTAAAGTTTAG 81321

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | | ||   | | |     |      | | ||| |       | |     | |  |  ||
Db      81320 AAAATTCAGTCAACTTTTACTGAGTTCTTACCAGGTCAAATTCTGGGAAAGACAAAATAT 81261

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
                | ||  |  ||        ||||  | |   |||  ||||    |  |||   | ||
Db      81260 AGT-TCCTATTTGGAAAAACTCATAATTTAATAGGACACATGACCTATAGCAAAATGTAA 81202

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              | |    ||  | ||    ||       |    |             |  |      | |
Db      81201 TCAGTTAATCTCATGTTAACAGATGGCAGGACTACTCCTCTCTTATTTACCTGGTCTTTG 81142

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTA-CAACTATTTAATAAAATTTTAAAAA 359
              |       || |    || ||  || |      | |||  |   || |||    |  ||
Db      81141 AAATAAACATGGCTCTTTAAACTGTAAAGGTGGACCAAACACAGAAAAAAGGAGTTGAAG 81082

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              ||    | | |   | ||   | |          |||||| |  || |||| ||||||
Db      81081 ATAGGTGAATGAGCCAAGTATGAAACCACCTCTCTAGCCAGGTGTGGTGGCTCATGCCTG 81022

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              ||||| ||||| |||||||||||  |||  || |||||||  |||||||||||||| |||||
Db      81021 TAATCTCAACAATTTGGGAGGCTGAGGCCGGTGGATCACTTGAGGTCAGGAGTTCAAGAC 80962

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              |||||||||||| |||||||||||| ||||||||||||  ||||||||||||||||||||
Db      80961 CAGCCTGGCCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCCAGGTGT 80902

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||| |   |||||||| | ||| | |||| ||||| ||||| ||||||||   ||||||
Db      80901 GGTAGTGGGCACCTGTAATGCCAGCCACTCGGGAGGCTGAGGCAGGAGAATCCTTTGAAC 80842

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAACA 655
              | |||||  |||| |||||| | ||||||   || || ||||||||||||| |||||||
Db      80841 CCAGGAGGCCGAGGCTGCAGTGAGCCGAGATTGTGCCATTGCACTCCAGCCTGGGCAACA 80782

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTAC 710
              |||||||||||  |    | ||  | |    | |||  |   |||| |  | | |
Db      80781 GAGCAAGACTCTGTCTCAAAAAATATAAAAATAAAAAAGGAGAAATCACATCTCC 80727


RESULT 18
ADE43315
ID   ADE43315 standard; DNA; 202100 BP.
XX
AC   ADE43315;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human IDE/ KNSL1 genomic sequence, SEQ ID 484.
XX
KW   Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;
KW   Alzheimer's disease; neuroprotective; nootropic; gene therapy;
KW   Chromosome 10; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003054143-A2.
XX
PD   03-JUL-2003.
XX
PF   25-OCT-2002; 2002WO-US034679.
XX
PR   25-OCT-2001; 2001US-0339525P.
PR   08-NOV-2001; 2001US-0336929P.
PR   08-NOV-2001; 2001US-0338010P.
PR   09-NOV-2001; 2001US-0338363P.
PR   04-DEC-2001; 2001US-0337052P.
PR   28-MAR-2002; 2002US-0368919P.
XX
PA   (NEUR-) NEUROGENETICS INC.
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Becker KD,  Velicelebi G,  Elliott KJ,  Wang X,  Tanzi RE,  Bertram L;
PI   Saunders AJ,  Mullin KM,  Sampson AJ,  Blacker DL;
XX
DR   WPI; 2003-559131/52.
XX
PT   Determining a predisposition for or the occurrence of neurodegenerative
PT   disease, e.g. Alzheimer's disease by detecting in a target nucleic acid
PT   the presence or absence of an allelic variant of one or more polymorphic
PT   regions.
XX
PS   Claim 9; Page 769-823; 848pp; English.
XX
CC   The present invention relates to a method (M1) for determining a
CC   predisposition for or the occurrence of neurodegenerative disease in a
CC   subject. The method comprises detecting in a target nucleic acid obtained
CC   from the subject the presence or absence of an allelic variant of one or
CC   more polymorphic regions of one or more genes selected from uPA
CC   (Urokinase plasminogen activator), SNCG (gamma-synuclein), IDE (insulin-
CC   degrading enzyme), KNSL1 (Kinesin-like protein 1), LIPA (lysosomal acid
CC   lypase), and TNFRSF6 (Tumour Necrosis Factor Receptor-SF6), where the
CC   presence of at least one of the allelic variant of one or more
CC   polymorphic regions is indicative of a predisposition for or the
CC   occurrence of neurodegenerative disease. The genes are all located on
CC   chromosome 10. M1 is useful for determining a predisposition for or the
CC   occurrence of, and for treating neurodegenerative disease, particularly
CC   Alzheimer's disease.
XX
SQ   Sequence 202100 BP; 60747 A; 41352 C; 41113 G; 58888 T; 0 U; 0 Other;

  Query Match            10.5%;  Score 210.6;  DB 10;  Length 202100;
  Best Local Similarity  60.6%;  Pred. No. 3.9e-34;
  Matches  433;  Conservative    0;  Mismatches  274;  Indels    8;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| ||||  ||||| ||||||||||||||||||||||||||||||||||||||
Db      42330 TGATCCACCTGCCTCAGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 42389

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | | || || ||| |||    |||| | ||      |   | ||  | |||  | || |
Db      42390 GCCTAAAGTGTTTCTTAAAGTAAGAAAAAAGTCAAG--AAAGATTAGGCTTAAAGTTTAG 42447

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | | ||   | | |     |      | | ||| |       | |     | |  |  ||
Db      42448 AAAATTCAGTCAACTTTTACTGAGTTCTTACCAGGTCAAATTCTGGGAAAGACAAAATAT 42507

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
                | ||  |  ||        ||||  | |   |||  ||||    |  |||   | ||
Db      42508 AGT-TCCTATTTGGAAAAACTCATAATTTAATAGGACACATGACCTATAGCAAAATGTAA 42566
```

```
Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              | |    ||   | ||    ||       |    |             |  |      | |
Db      42567 TCAGTTAATCTCATGTTAACAGATGGCAGGACTACTCCTCTCTTATTTACCTGGTCTTTG 42626

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTA-CAACTATTTAATAAAATTTTAAAAA 359
              |      || |    || ||  || |       | |||  |   || |||     |  ||
Db      42627 AAATAAACATGGCTCTTTAAACTGTAAAGGTGGACCAAACACAGAAAAAAGGAGTTGAAG 42686

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              ||    | | |    | ||   | |            |||||| |  || |||| ||||||
Db      42687 ATAGGTGAATGAGCCAAGTATGAAACCACCTCTCTAGCCAGGTGTGGTGGCTCATGCCTG 42746

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              ||||| ||||| ||||||||||  |||  || |||||||  |||||||||||||| |||||
Db      42747 TAATCTCAACAATTTGGGAGGCTGAGGCCGGTGGATCACTTGAGGTCAGGAGTTCAAGAC 42806

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              |||||||||||| |||||||||||||  |||||||||||||  |||||||||||||||||
Db      42807 CAGCCTGGCCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCCAGGTGT 42866

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              |||  |   |||||||| | ||| | |||| ||||| ||||| ||||||||   ||||||
Db      42867 GGTAGTGGGCACCTGTAATGCCAGCCACTCGGGAGGCTGAGGCAGGAGAATCCTTTGAAC 42926

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAACA 655
              | ||||||  |||| |||||| | ||||||   || || ||||||||||||| |||||||
Db      42927 CCAGGAGGCCGAGGCTGCAGTGAGCCGAGATTGTGCCATTGCACTCCAGCCTGGGCAACA 42986

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTAC 710
              |||||||||||  |    | ||  | |    | |||  |   |||| |  | | |
Db      42987 GAGCAAGACTCTGTCTCAAAAAATATAAAAATAAAAAAGGAGAAATCACATCTCC 43041
```

RESULT 19
ADH54357
ID ADH54357 standard; DNA; 202100 BP.
XX
AC ADH54357;
XX
DT 25-MAR-2004 (first entry)
XX
DE Human IDE/KNSL1 gene DNA sequence SeqID484.
XX
KW human; neurodegenerative disease; urokinase plasminogen activator; uPA;
KW gamma-synuclein; SNCG; insulin degrading enzyme; IDE;
KW kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;
KW tumour necrosis factor receptor SF6; TNFRSF6; Alzheimer's disease; ds.
XX
OS Homo sapiens.
XX
PN US2003224380-A1.
XX
PD 04-DEC-2003.
XX
PF 25-OCT-2002; 2002US-00282174.
XX
PR 25-OCT-2001; 2001US-0339525P.
PR 25-OCT-2001; 2001US-0348065P.
PR 02-NOV-2001; 2001US-0336983P.
PR 08-NOV-2001; 2001US-0336929P.
PR 08-NOV-2001; 2001US-0338010P.
PR 09-NOV-2001; 2001US-0338363P.
PR 04-DEC-2001; 2001US-0337052P.
PR 28-MAR-2002; 2002US-0368919P.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Becker KD, Velicelebi G, Ellliott KJ, Wang X, Tanzi RE;
PI Bertram L, Saunders AJ, Mullin KM, Sampson AJ;
XX
DR WPI; 2004-060538/06.
XX
PT Determining a predisposition for or the occurrence of neurodegenerative
PT disease, particularly Alzheimer's disease, comprises determining the
PT presence of a polymorphism in the uPA, SNCG, IDE, KNSL1, LIPA or TNFRSF6
PT gene.
XX
PS Claim 9; SEQ ID NO 484; 205pp; English.
XX
CC This invention relates to a novel method of determining a predisposition
CC for or the occurrence of neurodegenerative disease comprising detecting
CC in a target nucleic acid obtained from the subject the presence of an
CC allelic variant of polymorphic regions of human genes selected from
CC urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin
CC degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid
CC lipase (LIPA) and tumour necrosis factor receptor SF6 (TNFRSF6). The
CC method is useful in determining the presence or predisposition to a
CC neurodegenerative disease, particularly Alzheimer's disease. The present
CC sequence is that of a region of human DNA, containing the human IDE and
CC KNSL1 genes, which is related to the invention.
XX
SQ Sequence 202100 BP; 60747 A; 41352 C; 41113 G; 58888 T; 0 U; 0 Other;

Query Match 10.5%; Score 210.6; DB 12; Length 202100;
Best Local Similarity 60.6%; Pred. No. 3.9e-34;
Matches 433; Conservative 0; Mismatches 274; Indels 8; Gaps 5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||||| ||||  ||||| ||||||||||||||||||||||||||||||||||||||
Db      42330 TGATCCACCTGCCTCAGCCTCTCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 42389

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | | || || ||| |||    |||| | ||     |    | ||  | |||  | || |
Db      42390 GCCTAAAGTGTTTCTTAAAGTAAGAAAAAAGTCAAG--AAAGATTAGGCTTAAAGTTTAG 42447

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | | ||   |  | |   |      | | |||  |      | |    | |  |  ||
Db      42448 AAAATTCAGTCAACTTTTACTGAGTTCTTACCAGGTCAAATTCTGGGAAAGACAAAATAT 42507

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | ||  |  ||       ||||  | |   |||   ||||    |  |||   | ||
Db      42508 AGT-TCCTATTTGGAAAAACTCATAATTTAATAGGACACATGACCTATAGCAAAATGTAA 42566

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              | |    ||   | ||    ||       |    |             |  |      | |
Db      42567 TCAGTTAATCTCATGTTAACAGATGGCAGGACTACTCCTCTCTTATTTACCTGGTCTTTG 42626

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTA-CAACTATTTAATAAAATTTTAAAAA 359
              |      || |    || ||  || |       | |||  |   || |||     |  ||
Db      42627 AAATAAACATGGCTCTTTAAACTGTAAAGGTGGACCAAACACAGAAAAAAGGAGTTGAAG 42686

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              ||    | | |    | ||   | |            |||||| |  || |||| ||||||
Db      42687 ATAGGTGAATGAGCCAAGTATGAAACCACCTCTCTAGCCAGGTGTGGTGGCTCATGCCTG 42746

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              ||||| ||||| ||||||||||  |||  || |||||||  |||||||||||||| |||||
Db      42747 TAATCTCAACAATTTGGGAGGCTGAGGCCGGTGGATCACTTGAGGTCAGGAGTTCAAGAC 42806

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              |||||||||||| |||||||||||||  |||||||||||||  |||||||||||||||||
Db      42807 CAGCCTGGCCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCCAGGTGT 42866

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              |||  |   |||||||| | ||| | |||| ||||| ||||| ||||||||   ||||||
Db      42867 GGTAGTGGGCACCTGTAATGCCAGCCACTCGGGAGGCTGAGGCAGGAGAATCCTTTGAAC 42926

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCT-GGCAACA 655
              | ||||||  |||| |||||| | ||||||   || || ||||||||||||| |||||||
Db      42927 CCAGGAGGCCGAGGCTGCAGTGAGCCGAGATTGTGCCATTGCACTCCAGCCTGGGCAACA 42986

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTAC 710
              |||||||||||  |    | ||  | |    | |||  |   |||| |  | | |
Db      42987 GAGCAAGACTCTGTCTCAAAAAATATAAAAATAAAAAAGGAGAAATCACATCTCC 43041
```

```
RESULT 20
AEA61169_2/c
Continuation (3 of 4) of AEA61169 from base 200001 (Human CACNB2 gene genomic sequence S
WP Sequence split into 4 fragments  LOCUS AEA61169 Accession Aea61169
WP     Fragment Name      Begin     End
WP     AEA61169_0             1    110000
WP     AEA61169_1        100001    210000
WP     AEA61169_2        200001    310000
WP     AEA61169_3        300001    401433

  Query Match             10.5%;  Score 210.2;  DB 14;  Length 110000;
  Best Local Similarity   77.8%;  Pred. No. 4.3e-34;
  Matches  267;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              | |  | |  | ||| | ||| |      ||  ||||||||| || ||||||    ||||
Db      57128 TCAGGCCCACATATTTTAAAATAAAGCAAAATGGGGCCAGGCACAATAGCTCACAACTGT 57069

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||| ||| |||||||| |||||||||| ||||||||||   |||||
Db      57068 AATCCCAGCACTTTGGGAAGCCGAGGTGGGTGGATCACCTGTGGTCAGGAGTTCGAGACC 57009

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||| ||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db      57008 AGCCTGACCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGTGTG 56949

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||    | | || ||||| |||||| ||||| ||||| |||||||| |||||||||
Db      56948 GTGGCACATGCATATAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 56889

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGCAACAGA 657
              ||||||||||||||||||| | | ||||    || ||||||||||||||||||||||||
Db      56888 CTGGAGGTGGAGGTTGCAGTGAGCTGAGACTGTGCCACTGCACTCCAGCCTGGCAACAGA 56829

Qy        658 GCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAA 700
              || ||||||| |    |  |    |   |  |||| |  || |
Db      56828 GCGAGACTCCGTCTCAATCAATCATCAATCAAAATGGGATAGA 56786


RESULT 21
ABX16034/c
ID   ABX16034 standard; DNA; 203654 BP.
XX
AC   ABX16034;
XX
DT   04-APR-2003  (first entry)
XX
DE   Human gene encoding calcium channel transporter family member.
XX
KW   Human; ds; gene; chromosome 10; calcium channel transporter; cytostatic;
KW   liver; adrenal gland; tumorous nervous tissue; adult amygdala;
KW   brain meningioma; denis-drash; adult brain; foetal brain; placenta;
KW   testis; kidney; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(16802,G)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17103,A)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(71782,C)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(102639,G)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(132707,C)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(140698,C)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(140888,C)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(144278,Y)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(148851,C)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(159121,C)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(161499,G)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(161725,G)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(162859,T)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(162981,G)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(167468,R)
FT                   /*tag=  o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(167522,G)
FT                   /*tag=  p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(168351,G)
FT                   /*tag=  q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(168397,C)
FT                   /*tag=  r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(168534,T)
FT                   /*tag=  s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(168806,C)
FT                   /*tag=  t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(170017,T)
FT                   /*tag=  u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(171371,G)
FT                   /*tag=  v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(172828,T)
FT                   /*tag=  w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(174564,W)
FT                   /*tag=  x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(176912. .176914,AC)
FT                   /*tag=  y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(176912. .176914,AAC)
FT                   /*tag=  z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(182424,G)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(183891. .183893,AT)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(184320,A)
FT                   /*tag=  ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(184426,G)
FT                   /*tag=  ad
```

```
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(187085,G)
FT                   /*tag=  ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(187666,C)
FT                   /*tag=  af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(187729,G)
FT                   /*tag=  ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(187734,T)
FT                   /*tag=  ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(187739,T)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(187740,A)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(188292,C)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(188742,G)
FT                   /*tag=  al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(189033,T)
FT                   /*tag=  am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(192341,Y)
FT                   /*tag=  an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(192846,G)
FT                   /*tag=  ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(194004,C)
FT                   /*tag=  ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(194941,A)
FT                   /*tag=  aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(195301,C)
FT                   /*tag=  ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(200286. .200288,TT)
FT                   /*tag=  as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(200322. .200324,AA)
FT                   /*tag=  at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(200872. .200874,AC)
FT                   /*tag=  au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(200872. .200874,AAC)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(203298,T)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   US2002142938-A1.
XX
PD   03-OCT-2002.
XX
PF   30-MAR-2001; 2001US-00820905.
XX
PR   30-MAR-2001; 2001US-00820905.
XX
PA   (YANC/) YAN C.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Yan C,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-174139/17.
DR   P-PSDB; ABG74119.
XX
PT   New peptides related to calcium channel transporter subfamily, useful for
PT   treating disorders associated with abnormal expression of the transporter
PT   in liver, adrenal gland, normal and tumorous nervous tissues, placenta,
PT   or kidney.
XX
PS   Claim 22; Fig 3; 193pp; English.
XX
CC   The invention relates to an isolated peptide (a member of the calcium
CC   channel transporter family) appearing as ABG74119 (or an allelic variant,
CC   orthologue or fragment of at least 10 contiguous amino acids), which is
CC   encoded by a nucleic acid molecule that hybridises under stringent
CC   conditions to the opposite strand of the cDNA and genomic DNA appearing
CC   as ABX16033 and ABX16034 (including their complement or a sequence 80%
CC   similar to them). Also included are antibodies, vectors, host cells, gene
CC   chips, transgenic non-human animals and identification of modulators and
CC   binding agents. The peptides are useful in substantial and specific
CC   assays related to functional information of the peptide sequences, to
CC   raise antibodies or to elicit immune response, as reagents in assays to
```

```
  Query Match             10.5%;  Score 210.2;  DB 10;  Length 203654;
  Best Local Similarity   77.8%;  Pred. No. 4.7e-34;
  Matches  267;  Conservative    0;  Mismatches   73;  Indels    3;  Gaps    1;

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              | |  | |  | ||| | ||| |      ||  ||||||||| || ||||||    ||||
Db      57806 TCAGGCCCACATATTTTAAAATAAAGCAAAATGGGGCCAGGCACAATAGCTCACAACTGT 57747

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||| ||| ||||||| |||||||||| |||||||||||  |||||
Db      57746 AATCCCAGCACTTTGGGAAGCCGAGGTGGGTGGATCACCTGTGGTCAGGAGTTCGAGACC 57687

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||| ||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db      57686 AGCCTGACCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGTGTG 57627

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||    | | || ||||| |||||| ||||| ||||| |||||||| |||||||||
Db      57626 GTGGCACATGCATATAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 57567

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGCAACAGA 657
              |||||||||||||||||| | | ||||    || ||||||||||||||||||||||||||
Db      57566 CTGGAGGTGGAGGTTGCAGTGAGCTGAGACTGTGCCACTGCACTCCAGCCTGGCAACAGA 57507

Qy        658 GCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAA 700
              || ||||||| |   |  |   |    |  |||| |  || |
Db      57506 GCGAGACTCCGTCTCAATCAATCATCAATCAAAATGGGATAGA 57464
```

```
RESULT 22
ADQ20017
ID   ADQ20017 standard; DNA; 260160 BP.
XX
AC   ADQ20017;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 2837.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
```

```
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 2837; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 260160 BP; 68455 A; 56280 C; 57380 G; 78045 T; 0 U; 0 Other;

  Query Match             10.5%;  Score 210.2;  DB 12;  Length 260160;
  Best Local Similarity   74.4%;  Pred. No. 4.9e-34;
  Matches  279;  Conservative    0;  Mismatches   93;  Indels    3;  Gaps    1;

Qy        311 TGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTT 370
              ||   | |||  || ||| | ||  |  |  ||  |      ||| |||  ||
Db      93728 TGGGCACTTAGTAGATATCTGTTGAATATTCTTTTTTTTTAATTATAAAGGAACTACAGG 93787

Qy        371 AAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACA 430
              |||||| | |  |        ||||||  ||    || |||| ||||||||||||||| ||
Db      93788 GAATTATGAGAATGTAAAACTATGGGCTGGGTATGGTGGCTCATGCCTGTAATCCCAGCA 93847

Qy        431 CTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCA 490
              ||||||||||  |||||| |||||| |||||||||||| ||  |||||||||||||||||
Db      93848 CTTTGGGAGGCTGAGGTGGGTGGATCATCTGAGGTCAGGAATTCGAGACCAGCCTGGCCA 93907

Qy        491 ACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCA 550
              |||||||||||| |  || ||||||||||  |||||||||||||||||||||||||||
Db      93908 ACATGGTGAAACTCCTTCACTACTAAAAATACAAAAATTAGCCAGGTGTGGTGGCATGTG 93967

Qy        551 CCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGG 610
              | ||||||||||| || ||| ||||| ||||| |||||||| |||||||||  |||| | |
Db      93968 CTTGTAGTCCCAGCTTCTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAG 94027

Qy        611 AGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCC 667
              |||||||||| | || ||||   | ||||||||||||||||||| |||||| |||||||
Db      94028 AGGTTGCAGTGAGCCAAGATTGCGCCACTGCACTCCAGCCTGGCGACAGAGCGAGACTCC 94087

Qy        668 ATAAAGACAACAAAA 682
              |    | || ||||
Db      94088 GTCTCAAAAAAAAAA 94102


RESULT 23
ADQ97721/c
ID   ADQ97721 standard; DNA; 178024 BP.
XX
AC   ADQ97721;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer associated sequence HD10-033, SEQ ID 698.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 698; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 178024 BP; 46690 A; 39569 C; 40105 G; 51640 T; 0 U; 20 Other;

  Query Match             10.5%;  Score 209.2;  DB 12;  Length 178024;
  Best Local Similarity   77.0%;  Pred. No. 7.5e-34;
  Matches  282;  Conservative    0;  Mismatches   78;  Indels    6;  Gaps    2;

Qy        320 AAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTA 379
              || |   ||||  | |      ||||  ||  |  |   ||  | ||    ||| | |||
Db      50367 AAAATCAATTTGGTTCTGAATATTAAAGAAGCTGCACTTACAGAAACGTAGAAAGTTTTA 50308

Qy        380 AAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAG 439
              | | ||||    |  |||| ||||  || ||||  |||||||||||||| ||||||||||
Db      50307 TAAAACTTT---ACAGGCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG 50251

Qy        440 GCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGA 499
              ||  ||||||| |||||||||||  ||||||||| |||||||||||||||||||||||||
Db      50250 GCTGAGGTGGGTGGATCACCTGAGATCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGA 50191

Qy        500 AACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTC 559
              |||||||||||||||||||| |||||||||||| |  ||||||||  ||||||||||
Db      50190 AACCCTGTCTCTACTAAAAACACAAAAATTAGCCATGCATGGTGGCAGCTACCTGTAGTC 50131

Qy        560 CCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAG 619
              ||| |||||  ||||| ||| | |||||||| |||||||| ||||||||||||||||||
Db      50130 CCAGCTACTTGGGAGGCTGAAGCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGTTGCAG 50071

Qy        620 TAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACA 676
              | |   |||||   | || ||||||||||||||  ||||||||||||   | || | |
Db      50070 TGAGATGAGATGGCGCCATTGCACTCCAGCCTGGTGACAGAGCAAGACTGTCTCAAAAAA 50011

Qy        677 ACAAAA 682
              | ||||
Db      50010 AAAAAA 50005


RESULT 24
ADD71350/c
ID   ADD71350 standard; DNA; 93390 BP.
XX
AC   ADD71350;
XX
```

```
DT   15-JAN-2004  (first entry)
XX
DE   Glutamine:fructose-6-phosphate amide transferase 1 genomic sequence.
XX
KW   diabetes; haplotype; polymorphism; diagnosis; renopathy; intron; ds;
KW   glutamine:fructose-6-phosphate amide transferase 1.
XX
OS   Homo sapiens.
XX
PN   WO2003023063-A1.
XX
PD   20-MAR-2003.
XX
PF   06-SEP-2002; 2002WO-JP009093.
XX
PR   07-SEP-2001; 2001JP-00271870.
PR   28-MAR-2002; 2002JP-00090861.
XX
PA   (SANY ) SANKYO CO LTD.
XX
PI   Itakura M,  Yasumo H,  Watanabe I;
XX
DR   WPI; 2003-313261/30.
XX
PT   Judging relative onset risk of diabetes including type I or II diabetes
PT   and renopathy with or without type II diabetes accompanying, by detecting
PT   haplotype with gene polymorphism from human genomic DNA.
XX
PS   Claim 1; SEQ ID NO 22; 157pp; Japanese.
XX
CC   The invention relates to a method of judging the onset risk of diabetes
CC   comprising detecting a haplotype consisting of gene polymorphism at 1 or
CC   more positions selected from (a)-(h) from a specimen containing human
CC   genomic DNA supplied by a patient: (a) the nucleotide located at position
CC   36 of the intron 1 on GFAT1 (glutamine:fructose-6-phosphate amide
CC   transferase 1) gene (nucleotide number 632 in sequence ADD71329; (b) the
CC   nucleotide located at position 7 of the intron 11 on GFAT1 gene
CC   (nucleotide number 266 in sequence ADD71330; (c) the nucleotide located
CC   at position -147 of the intron 12 on GFAT1 gene (nucleotide number 338 in
CC   sequence ADD71331; (d) the nucleotide located at positions 1853-1877 of
CC   the intron 8 on GFAT1 gene (nucleotide numbers 336-360 in sequence
CC   ADD71332; (e) the nucleotide located at positions 1988-2007 of the intron
CC   12 on GFAT1 gene (nucleotide numbers 328-347 in sequence ADD71333; (f)
CC   the nucleotide located at position -11 to -22 of the intron 18 on GFAT1
CC   gene (nucleotide numbers 253-264 in sequence ADD71334; (g) the nucleotide
CC   located at positions 2632-2661 of the intron 3 on GFAT1 gene (nucleotide
CC   numbers 237-266 in sequence ADD71335; and (h) the nucleotide located at
CC   position 66 of the intron 18 on GFAT2 gene (nucleotide number 225 in
CC   sequence ADD71351). The method is useful for judging relative onset risk
CC   of diabetes including type I or II diabetes and renopathy with or without
CC   type II diabetes accompanying. This sequence represents the GFAT1 genomic
CC   sequence containing the polymorphisms shown above.
XX
SQ   Sequence 93390 BP; 26851 A; 17344 C; 19969 G; 29217 T; 0 U; 9 Other;

  Query Match             10.4%;  Score 208.8;  DB 10;  Length 93390;
  Best Local Similarity   75.7%;  Pred. No. 8.4e-34;
  Matches  286;  Conservative    0;  Mismatches   87;  Indels    5;  Gaps    2;

Qy        308 TATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACA 367
              | |  || |  ||| | |||   || |  |  |  ||| || | |  || ||  ||   |
Db      32191 TCTCAAAAAAATAAAAATTAAAAATAAATAAAAAATAAAAAGAATAAAACAAAGAAGTTA 32132

Qy        368 CTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCA 427
              |  | ||   | || ||       |||||||||| | |  ||||  ||||||||||||
Db      32131 CCAACATATATTAACAGAAGGCTGATGGGCCAGGCACGGCGGCTCACGCCTGTAATCCCA 32072

Qy        428 ACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGG 487
              |||||||||||||| ||| |||| ||||||  |||||||||| |  |||||| |||||
Db      32071 GCACTTTGGGAGGCCGAGGCGGGCAGATCAC--GAGGTCAGGAGATCGAGACCATCCTGG 32014

Qy        488 CCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCAT 547
              |||||||||||||||| |||||||||||||  ||||||||||  ||||||||||||
Db      32013 CCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCAG 31954

Qy        548 GCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGG 607
              |||||||| |||| |||||||||||| ||||| |||||||||||||||||   |||||
Db      31953 GCACCTGTAATCCCAGCTACTCAGGAGGGTGAGGCAGGAGAATTGCTTGAACTCGGGAGG 31894

Qy        608 TGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGAC 664
              |||||| |||||| | || ||||   | |||||||||||||||| |  ||||||  ||||
Db      31893 TGGAGGCTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTTGTGACAGAGTGAGAC 31834

Qy        665 TCCATAAAGACAACAAAA 682
              |||||    | || ||||
Db      31833 TCCATCTCAAAAAAAAAA 31816


RESULT 25
ADZ13446
ID   ADZ13446 standard; DNA; 243934 BP.
XX
AC   ADZ13446;
XX
DT   16-JUN-2005  (first entry)
XX
DE   Human cancer-associated genomic DNA #82.
XX
KW   Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
KW   cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2005031001-A2.
XX
PD   07-APR-2005.
XX
PF   23-SEP-2004; 2004WO-US031617.
XX
PR   23-SEP-2003; 2003US-00669920.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2005-273395/28.
XX
PT   Nucleic acid array useful for detecting cancer associated nucleic acid,
PT   comprises two or more nucleic acid probes.
XX
PS   Disclosure; SEQ ID NO 966; 198pp; English.
XX
CC   The invention relates to a nucleic acid array for detecting a cancer
CC   associated (CA) nucleic acid, comprising two or more nucleic acid probes.
CC   The invention also relates to a peptide array comprising two or more
CC   isolated polypeptides encoded by a CA nucleic acid sequence, a compound
CC   that binds to a polypeptide, an isolated antibody or its fragment which
CC   binds to a polypeptide, which is prepared by immunizing a host animal
CC   with a composition comprising the polypeptide or its antigen binding
CC   fragment and collecting cells from the host expressing antibodies against
CC   the antigen or its antigen binding fragment, a composition comprising the
CC   antibody and a carrier, a method of screening for anticancer activity, a
CC   method of detecting a CA nucleic acid, a method of diagnosing cancer, a
CC   method of treating cancer and a method of inhibiting expression of a CA
CC   nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
CC   nucleic acids. The antibody is useful for detecting the presence or
CC   absence of cancer cells in an individual which involves contacting cells
CC   from the individual with the antibody and detecting a complex of a CA
CC   protein from the cancer cells and the antibody, where the detection of
CC   the complex correlates with the presence of cancer cells in the
CC   individual. The composition is useful for inhibiting growth of cancer
CC   cells in an individual or for delivering a therapeutic agent to cancer
CC   cells in an individual. The invention is also useful for diagnosing
CC   cancer, for treating cancer and for inhibiting expression of a CA gene in
CC   a cell. This sequence represents human cancer-associated genomic DNA of
CC   the invention.
```

```
XX
SQ   Sequence 243934 BP; 66213 A; 49129 C; 53448 G; 74301 T; 0 U; 843 Other;

  Query Match            10.4%;  Score 208.8;  DB 14;  Length 243934;
  Best Local Similarity   77.0%;  Pred. No. 9.6e-34;
  Matches  268;  Conservative    0;  Mismatches   77;  Indels    3;  Gaps    1;

Qy        359 ACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCT 418
              | || |   |  | ||| || || |      || || ||| ||||| |  ||||   |||
Db      74796 AGTATTTGGCAGACATTTTTGAAAATGAACAAAGTGAGCCGGGCGCGGGGGCTCACACCT 74855

Qy        419 GTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGA 478
              ||||||||| ||||||||||||||||||||||| |||||||||||||| |||||| ||||
Db      74856 GTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCCGATCACCTGAGGTCCGGAGTTCAAGA 74915

Qy        479 CCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTG 538
              |||||||||||||||||||||||||| ||||||||||||||  |||||||||||  ||||
Db      74916 CCAGCCTGGCCAACATGGTGAAACCCAGTCTCTACTAAAAATACAAAAATTAGCTGGGTG 74975

Qy        539 TGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAA 598
              |||||||||||||||||||| || |||||||||||| ||||| ||||||||||||||||
Db      74976 TGGTGGCATGCACCTGTAGTCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAA 75035

Qy        599 CCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGCAACA 655
              ||  |||||||||||||||||| | | ||||   || || ||||||||||||||  |
Db      75036 CCCTGGAGGTGGAGGTTGCAGTGAACTGAGATGGTGCCATTGCACTCCAGCCTGGGCAAC 75095

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGA 703
               ||   ||  |   |     || ||||      ||||     || |||
Db      75096 AAGAGTGAAACTCCATTTCAAAAAAAAAAAAGAAAATGAACAAAGTGA 75143


RESULT 26
ADC86824/c
ID   ADC86824 standard; DNA; 17588 BP.
XX
AC   ADC86824;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1277.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC86825.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1277; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 17588 BP; 4287 A; 4481 C; 4353 G; 4267 T; 0 U; 200 Other;

  Query Match            10.4%;  Score 208.6;  DB 10;  Length 17588;
  Best Local Similarity   80.6%;  Pred. No. 7.3e-34;
  Matches  257;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    1;

Qy        390 AAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGG 449
              |||| |||  |||||||||||||   |||||||||||||||||||||||||||| ||| ||
Db      15131 AAATTGGCTGGGCGCAGTAGCTCACACCTGTAATCCCAACACTTTGGGAGGCCGAGGCGG 15072

Qy        450 GCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCT 509
              |||||| |||||||||||||||| |||||||| |||||||| ||||||||||||  |||
Db      15071 GCGGATTACCTGAGGTCAGGAGTTCAAGACCAGTCTGGCCAATATGGTGAAACCCCATCT 15012

Qy        510 CTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTC 569
              ||||||||||  ||||||||||| ||||||||||||  | ||||| ||||| ||||
Db      15011 CTACTAAAAATACAAAAATTAGCCGGGTGTGGTGGCACACGCCTGTAATCCCAGCTACAT 14952

Qy        570 AGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA 629
              ||||| ||||| | ||||||  |||||||| |||||| |||||||||||| | || |||
Db      14951 GGGAGGCTGAGGCAAGAGAATCACTTGAACCCAGGAGGAGGAGGTTGCAGTGAGCCAAGA 14892

Qy        630 T---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAAGCTT 686
              |   | |||||||||||||||||||||||||||||| |||||  |    | |  || |  |
Db      14891 TCTCGCCACTGCACTCCAGCCTGGCAACAGAGCCAGACTCTGTCTCAAAATAAATAAATA 14832

Qy        687 TGAAATTGTGTAAATGAGT 705
                || |   ||||| | |
Db      14831 AATAAATAAATAAATAAAT 14813


RESULT 27
ADO48536/c
ID   ADO48536 standard; DNA; 63266 BP.
XX
AC   ADO48536;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human neuropilin 1 (NRP1) genomic DNA sequence #2.
XX
KW   human; melanoma; single nucleotide polymorphism; SNP; neuropilin 1; NRP1;
KW   mannose receptor C type 2; MRC2; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       32929
FT                   /*tag=  a
FT                   /note= "Single nucleotide polymorphism"
FT   variation       34983
FT                   /*tag=  b
FT                   /note= "Single nucleotide polymorphism"
XX
PN   WO2004044163-A2.
XX
PD   27-MAY-2004.
XX
PF   06-NOV-2003; 2003WO-US035876.
XX
PR   06-NOV-2002; 2002US-0424475P.
PR   23-JUL-2003; 2003US-0489703P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Roth RB,  Nelson MR,  Braun A,  Kammerer SM;
XX
DR   WPI; 2004-411720/38.
```

```
XX
PT   Identifying a subject at risk of melanoma, useful for treating melanoma,
PT   comprises detecting the presence or absence of one or more polymorphic
PT   variations associated with melanoma in a nucleic acid sample from a
PT   subject.
XX
PS   Claim 23; SEQ ID NO 2; 176pp; English.
XX
CC   The invention comprises a method for identifying a subject at risk of
CC   melanoma. The invention involves detecting the presence or absence of one
CC   or more polymorphic variations associated with melanoma in the neuropilin
CC   1 (NRP1) or mannose receptor C type 2 (MRC2) genes. The method of the
CC   invention is useful for identifying subjects at risk and treating
CC   melanoma. The present nucleic acid represents the genomic DNA sequence
CC   for human NRP1.
XX
SQ   Sequence 63266 BP; 18633 A; 14176 C; 13204 G; 17251 T; 0 U; 2 Other;

  Query Match             10.4%;  Score 208.6;  DB 12;  Length 63266;
  Best Local Similarity   77.1%;  Pred. No. 8.7e-34;
  Matches  280;  Conservative    0;  Mismatches   79;  Indels    4;  Gaps    2;

Qy        326 TATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGC 385
              ||||||| ||||  | |   |      | |         | |  | |||  |  ||| | |
Db      34581 TATTTATCTCAACCTTGTTTCAGCTGATGAGTTCAGGCTGCTTTAAAAAAAAAAAAAATC 34522

Qy        386 TTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAG 445
                ||||  ||||||| ||| | |||| |||||||||||||| ||||||||||||||||
Db      34521 ACAGAAAGAGGCCAGGTGCAATGGCTCATGCCTGTAATCCCACCACTTTGGGAGGCCAAG 34462

Qy        446 GTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCT 505
              ||||| ||||||||||||||||||||||| ||||||||||||||||||||||||||| |||
Db      34461 GTGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAAACCT 34402

Qy        506 GTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACT 565
              || |||||||||||  |||||||||||  |||||||||||| || ||||   ||||| ||
Db      34401 GTTTCTACTAAAAATACAAAAATTAGCAGGGTGTGGTGGCACGCTCCTGCGATCCCAGCT 34342

Qy        566 ACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCC 625
              |||  ||||| ||||| ||||||||    ||||||| ||| ||||||||||||| | ||
Db      34341 ACTTGGGAGGCTGAGGCAGGAGAATCATCTGAACCTGGGAAGTGGAGGTTGCAGTGAGCC 34282

Qy        626 GAGA---TGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAA 681
              ||||   || || || |||||||||| ||| ||| |||||||||||||    | ||||||
Db      34281 GAGATTGTGCCATTGTACTCCAGCCTGGGCGACAAAGCAAGACTCCATCTCAAAAACAAA 34222

Qy        682 AGC 684
              | ||
Db      34221 ACC 34219


RESULT 28
ADX80721/c
ID   ADX80721 standard; DNA; 63266 BP.
XX
AC   ADX80721;
XX
DT   05-MAY-2005  (first entry)
XX
DE   Human nidogen 2 (NID2) genomic DNA.
XX
KW   melanoma; DNA polymorphism; SNP detection; cytostatic; gene therapy; SNP;
KW   single nucleotide polymorphism; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       32929
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       34983
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   WO2005017176-A2.
XX
PD   24-FEB-2005.
XX
PF   05-MAY-2004; 2004WO-US014238.
XX
PR   23-JUL-2003; 2003US-0489703P.
PR   06-NOV-2003; 2003US-00703789.
PR   06-NOV-2003; 2003US-00703817.
PR   06-NOV-2003; 2003US-00704513.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Roth RB,  Nelson MR,  Kammerer SM,  Braun A,  Hoyal-Wrightson CR;
XX
DR   WPI; 2005-182387/19.
DR   P-PSDB; ADX80738.
XX
PT   Identifying a subject at risk of melanoma by detecting presence or
PT   absence of a polymorphic variation associated with melanoma, where the
PT   presence of polymorphic variations is indicative of the subject being at
PT   risk of melanoma.
XX
PS   Example 21; SEQ ID NO 2; 418pp; English.
XX
CC   The invention relates to a novel method for identifying a subject at risk
CC   of melanoma. The method comprises detecting the presence or absence of a
CC   polymorphic variation associated with melanoma, where the presence of the
CC   one or more polymorphic variations is indicative of the subject being at
CC   risk of melanoma. The invention further comprises: a method for
CC   identifying a polymorphic variation associated with melanoma proximal to
CC   an incident polymorphic variation associated with melanoma; an isolated
CC   nucleic acid which comprises a portion of or all of a nucleotide sequence
CC   comprising fully defined 68400-213300 base pairs sequences (SEQ ID NO. 3,
CC   4, 5, 6, and/or 7) given in the specification, and comprises one or more
CC   polymorphic variations; an oligonucleotide comprising a nucleotide
CC   sequence complementary to a portion of the nucleotide sequence above,
CC   where the 3' end of the oligonucleotide is adjacent to a polymorphic
CC   variation; a microarray comprising the isolated nucleic acid linked to a
CC   solid support; an isolated polypeptide encoded by the isolated nucleic
CC   acid sequence; genotyping a nucleic acid; a method for identifying a
CC   candidate molecule that modulates cell proliferation; treating melanoma
CC   in a subject; and treating melanoma in a subject or preventing melanoma
CC   in a subject. The methods and sequences have cytostatic activity. The
CC   polynucleotides may be used in gene therapy. The methods are useful for
CC   identifying a subject at risk of melanoma, treating melanoma in a
CC   subject, or preventing melanoma in a subject. This polynucleotide
CC   sequence represents the polymorphic variation containing human nidogen 2
CC   (NID2) genomic DNA of the invention.
XX
SQ   Sequence 63266 BP; 18633 A; 14176 C; 13204 G; 17251 T; 0 U; 2 Other;

  Query Match             10.4%;  Score 208.6;  DB 14;  Length 63266;
  Best Local Similarity   77.1%;  Pred. No. 8.7e-34;
  Matches  280;  Conservative    0;  Mismatches   79;  Indels    4;  Gaps    2;

Qy        326 TATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGC 385
              ||||||| ||||  | |   |      | |         | |  | |||  |  ||| | |
Db      34581 TATTTATCTCAACCTTGTTTCAGCTGATGAGTTCAGGCTGCTTTAAAAAAAAAAAAAATC 34522

Qy        386 TTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAG 445
                ||||  ||||||| ||| | |||| |||||||||||||| ||||||||||||||||
Db      34521 ACAGAAAGAGGCCAGGTGCAATGGCTCATGCCTGTAATCCCACCACTTTGGGAGGCCAAG 34462

Qy        446 GTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCT 505
              ||||| ||||||||||||||||||||||| ||||||||||||||||||||||||||| |||
Db      34461 GTGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAAACCT 34402

Qy        506 GTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACT 565
              || |||||||||||  |||||||||||  |||||||||||| || ||||   ||||| ||
Db      34401 GTTTCTACTAAAAATACAAAAATTAGCAGGGTGTGGTGGCACGCTCCTGCGATCCCAGCT 34342
```

```
Qy        566 ACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCC 625
              ||| ||||| ||||| ||||||||     ||||||| ||| ||||||||||||| | ||
Db      34341 ACTTGGGAGGCTGAGGCAGGAGAATCATCTGAACCTGGGAAGTGGAGGTTGCAGTGAGCC 34282

Qy        626 GAGA---TGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAA 681
              ||||   || || || |||||||||| ||| ||| ||||||||||||    | ||||||
Db      34281 GAGATTGTGCCATTGTACTCCAGCCTGGGCGACAAAGCAAGACTCCATCTCAAAAACAAA 34222

Qy        682 AGC 684
              | |
Db      34221 ACC 34219


RESULT 29
ADQ17329
ID   ADQ17329 standard; DNA; 101685 BP.
XX
AC   ADQ17329;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 146.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 146; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 101685 BP; 27434 A; 22468 C; 23225 G; 28558 T; 0 U; 0 Other;

  Query Match             10.4%;  Score 208.6;  DB 12;  Length 101685;
  Best Local Similarity   58.8%;  Pred. No. 9.3e-34;
  Matches  433;  Conservative    0;  Mismatches  274;  Indels   30;  Gaps    3;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||| ||||||||||||||||||||||||||| |||||||| | ||
Db      51671 TGATCCACCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCACCCT 51730

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |      | |||   |||     ||   | | |  ||| |   | | ||  |    |
Db      51731 GCCAGGGATTCTTAAACTTTAGTAAGTATAAAAGTCCCTCGAGAGTTTGTTAAA----AG 51786

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | |  ||  |||  |  | |||||  | | |     |     |   |  | | | ||
Db      51787 ACAGATTCCTGGGCTTCATTCCCAGAGATTCTGACTTAGTAGGTGTTTCCAGTATCTACT 51846

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              ||| |        ||   ||  | ||     |    |         |||| |  ||
Db      51847 TCTAACAAGTACCCCTAGGTGATTTTGGCGTATTCCTTGGACCACACTTTGA--GATACC 51904

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |  |||  ||  |  |  | ||        |||  | | |  |  |    |  |
Db      51905 TGGACCGGTGGGAGATAGTAAGACATTAATTTACTTTTATTTGTGTGCATACAATCTGCT 51964

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||  | |     |  || |    | |||| || |  | |||||    ||||  ||||
Db      51965 TACTTTCATAAATAACTTCAGATAGTGTTTAATAAATATTTTTAAATTAATTAAAAAATT 52024

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              ||  |                          |   |||||||  |||||||   |||||
Db      52025 TAGGA-----------------------TTAGCAGCCAGGCGTGGTAGCTCACACCTGT 52060

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||| | |||||||||||||| |||| ||| ||||||||||||||||||||  |||||
Db      52061 AATCCTAGCACTTTGGGAGGCCCAGGTAGGCAGATCACCTGAGGTCAGGAGTTCGAGACC 52120

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||||||||| |||||||| |||||||||||||  ||||||||||  ||||||
Db      52121 AGCCTGGCCAACATGATGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTG 52180

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||| ||  | |||| ||||| |||| |||||||||||||| |||||||| ||| |||||
Db      52181 GTGGCGTGTGCATGTAATCCCAGCTACCCAGGAGGTTGAGGCAGGAGAATCGCTGGAACC 52240

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              |||||  ||||| |||||| | | |||||  |||| |    |   |  ||||||||||||
Db      52241 CGGGAGGCGGAGGCTGCAGTGAGCAGAGATCGCACTACTGCCAGCCTAGGCAACAGAGCA 52300

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              |||||||||    | || ||||      |||     |  | ||   | |   ||    ||
Db      52301 AGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAAAATTTAGGATCAGGAAAGATAGAGAT 52360

Qy        721 TTAAGAAATTCATCTTT 737
                |||  |   | | | |
Db      52361 AGAAGGATACCCTGTGT 52377


RESULT 30
AAL04768
ID   AAL04768 standard; DNA; 8553 BP.
XX
AC   AAL04768;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 7456.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
```

PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 7456; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a

```
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention
XX
SQ   Sequence 8553 BP; 2252 A; 1971 C; 1958 G; 2372 T; 0 U; 0 Other;

  Query Match             10.4%;  Score 208.4;  DB 4;  Length 8553;
  Best Local Similarity   84.8%;  Pred. No. 7.2e-34;
  Matches  246;  Conservative    0;  Mismatches   41;  Indels    3;  Gaps    1;

Qy        396 GCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGAT 455
              ||| ||||| || ||||  |||||||||||||| |||||||||||||| ||| ||| ||||
Db        541 GCCGGGCGCTGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGAT 600

Qy        456 CACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTA 515
              ||||||||||||||||  |||||||| |||||||||||| ||||||| ||||||||||
Db        601 CACCTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGGCGAAACCCCGTCTCTACTA 660

Qy        516 AAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGG 575
              ||||  |||||||||||| || ||||||||  || |||||| ||||| ||||||||||||
Db        661 AAAATACAAAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAATCCCAGCTACTCAGGAGG 720

Qy        576 TTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGT 632
               ||||| |||||||| |||||||||  ||||  |||||||||||| | |||| |   |||
Db        721 CTGAGGCAGGAGAATCGCTTGAACCCTGGAGACGGAGGTTGCAGTGAGCCGACATCGTGT 780

Qy        633 CACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||||||||||||||| ||||||| |||||||||     | || ||||
Db        781 CACTGCACTCCAGCCTGGCGACAGAGCGAGACTCCATCTCAAAAAAAAAA 830


RESULT 31
ABL97671
ID   ABL97671 standard; DNA; 8553 BP.
XX
AC   ABL97671;
XX
DT   21-JUN-2002  (first entry)
XX
DE   Human testicular antigen encoding DNA fragment SEQ ID NO: 2323.
XX
KW   Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW   reproductive system disorder; urinary system disorder; gene therapy;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155317-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001329.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
```

```
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483232/52.
XX
PT   Nucleic acids encoding 973 human testicular antigen polypeptides, useful
PT   for preventing, diagnosing and/or treating testicular cancer.
XX
PS   Disclosure; SEQ ID NO 2323; 766pp; English.
XX
CC   The present invention provides the protein and coding sequences of 973
CC   human testicular antigens, and fragments of their genomic sequences. The
CC   sequences can be used in the treatment of cardiovascular, urinary system,
CC   reproductive system, immune, respiratory, neurological and
CC   gastrointestinal disorders, infections, and particularly cancer,
CC   especially testicular cancers. The present sequence is a DNA encoding a
CC   protein fragment of the invention
XX
SQ   Sequence 8553 BP; 2252 A; 1971 C; 1958 G; 2372 T; 0 U; 0 Other;

  Query Match            10.4%;  Score 208.4;  DB 4;  Length 8553;
  Best Local Similarity   84.8%;  Pred. No. 7.2e-34;
  Matches  246;  Conservative    0;  Mismatches   41;  Indels    3;  Gaps    1;

Qy        396 GCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGAT 455
              ||| ||||| || ||||  |||||||||||||| |||||||||||||| ||| ||| ||||

Db        541 GCCGGGCGCTGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGAT 600

Qy        456 CACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTA 515
              |||||||||||||||||  |||||||| |||||||||||| ||||||| ||||||||||

Db        601 CACCTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGGCGAAACCCCGTCTCTACTA 660

Qy        516 AAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGG 575
              ||||  |||||||||||| || ||||||||  || |||||| ||||| ||||||||||||

Db        661 AAAATACAAAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAATCCCAGCTACTCAGGAGG 720

Qy        576 TTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGT 632
              ||||| |||||||| |||||||||  ||||  |||||||||||| | |||| |   |||

Db        721 CTGAGGCAGGAGAATCGCTTGAACCCTGGAGACGGAGGTTGCAGTGAGCCGACATCGTGT 780

Qy        633 CACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||||||||||||||| ||||||| |||||||||    | || ||||

Db        781 CACTGCACTCCAGCCTGGCGACAGAGCGAGACTCCATCTCAAAAAAAAAA 830


RESULT 32
ACN44462
ID   ACN44462 standard; DNA; 51928 BP.
XX
AC   ACN44462;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG38450.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 922; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 51928 BP; 12432 A; 9654 C; 9978 G; 13530 T; 0 U; 6334 Other;

  Query Match            10.4%;  Score 208.4;  DB 11;  Length 51928;
  Best Local Similarity   81.2%;  Pred. No. 9.3e-34;
  Matches  255;  Conservative    0;  Mismatches   56;  Indels    3;  Gaps    1;
```

```
Qy        372 AATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACAC 431
              ||| |  ||| |  ||||   | ||||||||||| || ||||  || ||||||||||| |||
Db      27729 AATGAAAAAAAATATTTGCCCTTGGCCAGGCGCGGTGGCTCACGCTTGTAATCCCAGCAC 27788

Qy        432 TTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAA 491
              |||||||||| |||||||| |||||||||||||| |||||||| ||||||||||||||||
Db      27789 TTTGGGAGGCAAAGGTGGGTGGATCACCTGAGGTTAGGAGTTTGAGACCAGCCTGGCCAA 27848

Qy        492 CATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCAC 551
              ||||||||||||  | ||||||||||  ||||||||||  |||||||||| ||||||
Db      27849 CATGGTGAAACCCCATATCTACTAAAAATACAAAAATTAGCGGGGTGTGGTGGTATGCAC 27908

Qy        552 CTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGA 611
              ||||| ||||| |||||| ||||| |||||  ||||||||||||||||| |||||| |||
Db      27909 CTGTAATCCCACCTACTCCGGAGGCTGAGGCTGGAGAATTGCTTGAACCCAGGAGGGGGA 27968

Qy        612 GGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCA 668
              |||||||||  | | |||||   | || ||||||||||||||  |  |||  | |||  |
Db      27969 GGTTGCAGTGAGCTGAGATCACGCCATTGCACTCCAGCCTGGGTAACGAGAGAAACTGTA 28028

Qy        669 TAAAGACAACAAAA 682
              |    | ||  ||||
Db      28029 TTTCAAAAAAAAAA 28042
```

RESULT 33
ADZ12979
ID ADZ12979 standard; DNA; 52124 BP.
XX
AC ADZ12979;
XX
DT 16-JUN-2005 (first entry)
XX
DE Human cancer-associated genomic DNA #44.
XX
KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
KW cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2005031001-A2.
XX
PD 07-APR-2005.
XX
PF 23-SEP-2004; 2004WO-US031617.
XX
PR 23-SEP-2003; 2003US-00669920.
XX
PA (CHIR) CHIRON CORP.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2005-273395/28.
XX
PT Nucleic acid array useful for detecting cancer associated nucleic acid,
PT comprises two or more nucleic acid probes.
XX
PS Disclosure; SEQ ID NO 499; 198pp; English.
XX
CC The invention relates to a nucleic acid array for detecting a cancer
CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.
CC The invention also relates to a peptide array comprising two or more
CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound
CC that binds to a polypeptide, an isolated antibody or its fragment which
CC binds to a polypeptide, which is prepared by immunizing a host animal
CC with a composition comprising the polypeptide or its antigen binding
CC fragment and collecting cells from the host expressing antibodies against
CC the antigen or its antigen binding fragment, a composition comprising the
CC antibody and a carrier, a method of screening for anticancer activity, a
CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a
CC method of treating cancer and a method of inhibiting expression of a CA
CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
CC nucleic acids. The antibody is useful for detecting the presence or
CC absence of cancer cells in an individual which involves contacting cells
CC from the individual with the antibody and detecting a complex of a CA
CC protein from the cancer cells and the antibody, where the detection of
CC the complex correlates with the presence of cancer cells in the
CC individual. The composition is useful for inhibiting growth of cancer
CC cells in an individual or for delivering a therapeutic agent to cancer
CC cells in an individual. The invention is also useful for diagnosing
CC cancer, for treating cancer and for inhibiting expression of a CA gene in
CC a cell. This sequence represents human cancer-associated genomic DNA of
CC the invention.
XX
SQ Sequence 52124 BP; 14171 A; 11009 C; 11507 G; 15437 T; 0 U; 0 Other;

```
  Query Match             10.4%;  Score 208.4;  DB 14;  Length 52124;
  Best Local Similarity   81.2%;  Pred. No. 9.3e-34;
  Matches  255;  Conservative    0;  Mismatches   56;  Indels    3;  Gaps    1;

Qy        372 AATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACAC 431
              |||  |  |||  |  ||||   | |||||||||| || ||||  || ||||||||||  |||
Db      27728 AATGAAAAAAAATATTTGCCCTTGGCCAGGCGCGGTGGCTCACGCTTGTAATCCCAGCAC 27787

Qy        432 TTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAA 491
              |||||||||| |||||||| |||||||||||||| |||||||| ||||||||||||||||
Db      27788 TTTGGGAGGCAAAGGTGGGTGGATCACCTGAGGTTAGGAGTTTGAGACCAGCCTGGCCAA 27847

Qy        492 CATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCAC 551
              ||||||||||||  | ||||||||||  ||||||||||  |||||||||| ||||||
Db      27848 CATGGTGAAACCCCATATCTACTAAAAATACAAAAATTAGCGGGGTGTGGTGGTATGCAC 27907

Qy        552 CTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGA 611
              ||||| ||||| |||||| ||||| |||||  |||||||||||||||| |||||| |||
Db      27908 CTGTAATCCCACCTACTCCGGAGGCTGAGGCTGGAGAATTGCTTGAACCCAGGAGGGGGA 27967

Qy        612 GGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCA 668
              |||||||||  | | |||||   | || ||||||||||||||  |  |||  | |||  |
Db      27968 GGTTGCAGTGAGCTGAGATCACGCCATTGCACTCCAGCCTGGGTAACGAGAGAAACTGTA 28027

Qy        669 TAAAGACAACAAAA 682
              |    | ||  ||||
Db      28028 TTTCAAAAAAAAAA 28041
```

RESULT 34
AAL05708/c
ID AAL05708 standard; DNA; 27154 BP.
XX
AC AAL05708;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 8396.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 8396; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 27154 BP; 7884 A; 5528 C; 6156 G; 7586 T; 0 U; 0 Other;

```
  Query Match             10.4%;  Score 208.2;  DB 4;  Length 27154;
  Best Local Similarity   60.3%;  Pred. No. 9.4e-34;
  Matches  419;  Conservative    0;  Mismatches  263;  Indels   13;  Gaps    4;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||||||||||||||||| ||||| || ||||||||||||||||||||||  | ||||
Db       1766 TGATCCACCAGCCTTGGCCTTCCAAAATGGTGGGATTACAGGCGTGAGCCACTGCACCTG 1707

Qy         61 GTCGAATGTC-TTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | |  ||  | || | | ||   | | |   ||||  |   | | | | |       |
Db       1706 GCCTCATTCCATTCAATTTTAACAAACATTTCATGAACTAAATACCAGACACTGTACTAG 1647

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                  | ||   ||||   |       |   |       | ||   | |
Db       1646 GCTGGGGATAGAGAAAGAAATGAAATACAGTTCTTGTCTCTGTGGAACTCCCAGGCTGAT 1587

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              |   |    ||   | |       ||    ||||   || |||         ||   | |
Db       1586 GAGGCAGATAGGCCTGTAAACAAATGAAGAAGGAAAAACAATCTAATAGAAAAATTGGCA 1527

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              | |  |  | | |     || |  | |     | |  ||    |   || | || ||
Db       1526 AAAGGCACAAGTAGGCAATTCAAAAAAAAAAATCAACTGGCTGATAAATGTGAAAAATGC 1467

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
                | |  |  | | |  ||| |  | ||     |  |     ||||||  |   | |
Db       1466 TCAACATCACTAGTAATTTTTATTGATAACCTCTTTATTGTATTAATACCAAACTGATGG 1407

Qy        360 CTAATACACTTAA--------ATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCT 411
                |  ||  | |        | |||  || |  ||| || | ||||||| || || |||
Db       1406 TAACCACCTTGTAAGTGAATTAATATGTAAAATATTTCAAGTAGGCCAGGTGCCGTGGCT 1347

Qy        412 CCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAG 471
              |  |||||||||||| |||||||||| |   ||| ||| ||||||||||||||||||||
Db       1346 CACGCCTGTAATCCCAGCACTTTGGGAAGTTGAGGCGGGAGGATCACCTGAGGTCAGGAG 1287

Qy        472 TTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAG 531
              ||  |||||||||| |||||||||| ||||| |||||||||||||  ||||||||||
Db       1286 TTCGAGACCAGCCTGCCCAACATGGTGGAACCCCGTCTCTACTAAAAATACAAAAATTAG 1227

Qy        532 CCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATT 591
              || || |||||||   |||||||||||||| |||||| ||||| |||||  |||||||||
Db       1226 CCGGGCGTGGTGGTGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGGAGAATT 1167

Qy        592 GCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT 648
              ||||||||  |||||  |||||||||||| | |||||   ||| ||||| | ||||||||
Db       1166 TCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGATCATGCCACTGTATTCCAGCCT 1107

Qy        649 -GGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
               ||| ||||||  ||||||  |     | || ||||
Db       1106 AGGCGACAGAGTGAGACTCTGTCTCAAAAAAAAA 1072


RESULT 35
ADQ97557
ID   ADQ97557 standard; DNA; 254396 BP.
XX
AC   ADQ97557;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer associated sequence HD09-013, SEQ ID 534.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 534; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 254396 BP; 79338 A; 46516 C; 47859 G; 80683 T; 0 U; 0 Other;

  Query Match             10.4%;  Score 208.2;  DB 12;  Length 254396;
  Best Local Similarity   59.5%;  Pred. No. 1.3e-33;
  Matches  411;  Conservative    0;  Mismatches  268;  Indels   12;  Gaps    3;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||| ||    ||  |||||||||| |||| ||  ||| | |    ||||||| ||
Db     231739 TGATCCTCCCAACTCAGCCTCCCAAATAGCTGAGACCACATGTGCACACCACCACACCCA 231798

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              |    | |||||||| |||| |||||||   |    ||  | |       ||  |
Db     231799 GCTATTTTTCTTTATTTTTTGTAGAGACAGGGTCTCACTGCATTGCTCAGGCTGGTCTTG 231858

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              |     ||          |||||      |      ||        |    |     || |
Db     231859 AACTCCTGGGCTCAAGCAATCCTCCCACCATGGCTATAATCTCAAAGTGCTGGGATTATA 231918

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
                ||          |      || |  |||   |    ||| |         | |
Db     231919 GTCATCAGCCACTGCACCTGGCCTGGGTATCCTACTTTAAAGGAAATACAAATTGCTATG 231978

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGAT------TTCAAC 294
              |||   |  ||||  | ||   |  ||||   | ||    ||| |      |   |
Db     231979 AAAATTCAGAGAAAGAATAGACACTGTGGTTTTGGAGGACTCTAGGTTGCCTTGTCAGAA 232038

Qy        295 AAGTAGTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTT 354
              | | |||| ||  |||    | | |  ||  |||   || | | || | |   |
Db     232039 AGGAAGTATTTGAGCTTGGCTTTAAAGAATGCATGGATTTTGACAAGTGAAGATGAGCAT 232098

Qy        355 AAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCT 414
                | | | | | ||      ||  |  | ||| || ||||||||| || ||||
Db     232099 T---ATTTAGAAATTCCAAGAGCAAAGAAAATATGATATTGGCCAGGCGCGGTGGCTCAC 232155

Qy        415 GCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTT 474
              ||||||| |||||||||||||||||||| |||||||||  |||||||||  |||||||||
Db     232156 GCCTGTACTCCCAACACTTTGGGAGGCCGAGGTGGGCGGCTCACCTGAGATCAGGAGTTT 232215

Qy        475 AAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCA 534
              ||||||  |||||||||||||||||||||   |||||||||||||  ||||||||||||
Db     232216 GAGACCATCCTGGCCAACATGGTGAAACCTCATCTCTACTAAAAATACAAAAATTAGCCA 232275

Qy        535 GGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCT 594
              |||||||||||| |||||  ||| |||||  ||||||||||||  ||||  | | ||||  ||
Db     232276 GGTGTGGTGGCAGGCACCCGTAATCCCAGCTACTCAGGAGGCTGAGACAAGTGAATCACT 232335

Qy        595 TGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGC 651
              |||||| |||||| |||||||||| | | ||||||    || |||||||| |||||||
Db     232336 TGAACCCAGGAGGCGGAGGTTGCATTGAGCCGAGATTGTGCCACTGCACTGCAGCCTGTG 232395

Qy        652 AACAGAGCAAGACTCCATAAAGACAACAAAA 682
                  ||||| ||| ||  |    | || ||||
```

```
Db     232396 TGAAGAGCGAGATTCTGTCTCAAAAAAAAAA 232426


RESULT 36
AAK68575
ID   AAK68575 standard; DNA; 32145 BP.
XX
AC   AAK68575;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23387.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
```

```
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 23387; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 32145 BP; 8777 A; 7496 C; 7117 G; 8755 T; 0 U; 0 Other;

  Query Match             10.4%;  Score 208;  DB 4;  Length 32145;
  Best Local Similarity   83.8%;  Pred. No. 1.1e-33;
  Matches  248;  Conservative    0;  Mismatches   45;  Indels    3;  Gaps    1;

Qy        390 AAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGG 449
              | |  ||||||| || || ||||   |||| ||||||| ||||||||||||||||||||
Db      25596 ATAGAGGCCAGGTGCGGTGGCTCACACCTGCAATCCCAGCACTTTGGGAGGCCAAGGTGG 25655

Qy        450 GCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCT 509
              || ||||||||||||||||||||  |||||||||||||||||||||||||| |||||||
Db      25656 GCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAAGCCTGTCT 25715

Qy        510 CTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTC 569
              |||||||||| |||||||||||| ||  ||||||   |||||||||  |||||| ||||||
Db      25716 CTACTAAAAATGCAAAAATTAGCCGGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTC 25775

Qy        570 AGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG- 628
              |||||| ||||| |||||||||||||||||||| ||||| || ||||||||| | |||||
Db      25776 AGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGTGGTTGCAGTGAGCCGAGA 25835

Qy        629 --ATGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
                ||| ||||| |||||||||||||| ||||| ||| ||  |      || ||||
Db      25836 TCATGCCACTGTACTCCAGCCTGGCAATAGAGCTAGATTCTGTCTCAGAAAAAAAA 25891


RESULT 37
AAK68491
ID   AAK68491 standard; DNA; 32145 BP.
XX
AC   AAK68491;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23303.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
```

PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 23303; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 32145 BP; 8777 A; 7496 C; 7117 G; 8755 T; 0 U; 0 Other;

Query Match 10.4%; Score 208; DB 4; Length 32145;
Best Local Similarity 83.8%; Pred. No. 1.1e-33;
Matches 248; Conservative 0; Mismatches 45; Indels 3; Gaps 1;

```
Qy      390 AAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGG 449
            | |  ||||||| || || ||||   |||| ||||||| ||||||||||||||||||||
Db    25596 ATAGAGGCCAGGTGCGGTGGCTCACACCTGCAATCCCAGCACTTTGGGAGGCCAAGGTGG 25655

Qy      450 GCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCT 509
            || |||||||||||||||||||||  |||||||||||||||||||||||||| |||||||
Db    25656 GCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAAGCCTGTCT 25715

Qy      510 CTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTC 569
            |||||||||| |||||||||||| ||  ||||||   ||||||||| ||||| ||||||
Db    25716 CTACTAAAAATGCAAAAATTAGCCGGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTC 25775

Qy      570 AGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG- 628
            |||||| ||||| |||||||||||||||||||| ||||| || |||||||| | |||||
Db    25776 AGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGTGGTTGCAGTGAGCCGAGA 25835

Qy      629 --ATGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              ||| ||||| |||||||||||||| ||||| ||| ||  |       || ||||
Db    25836 TCATGCCACTGTACTCCAGCCTGGCAATAGAGCTAGATTCTGTCTCAGAAAAAAAA 25891
```

RESULT 38
AAD58977
ID AAD58977 standard; DNA; 83450 BP.
XX
AC AAD58977;
XX
DT 18-DEC-2003 (first entry)

```
XX
DE   Human phosphatase genomic DNA.
XX
KW   Human; phosphatase protein; therapeutic; diagnostic; immune response;
KW   pharmacogenomic; tissue typing; gene therapy; chromosome 15; enzyme;
KW   cancer; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(85,T)
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1614, G)
FT                   /*tag= b
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(2648, G)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(6773, G)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(8144, A)
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9488, C)
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9730, G)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10317, C)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10391, C)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10682, A)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(11632. .11634, TT)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12054, A)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12938, G)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13719, A)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14332. .14334, TT)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14753, G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14830. .14832, AT)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14831, A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15157, A)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16242, A)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16698, G)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17152, G)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17623. .17625, TA)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17624, T)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18267. .18269, AA)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18786, G)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18812. .18814, CG)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18813, A)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22061, G)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22205, T)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22527, A)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22876, G)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(23351, A)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(23821, G)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(25487, T)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(30858, C)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(31014, C)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(31437, T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35341, G, A)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35346, G)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36164, A)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36620, G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37074, G)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37545. .37547, TA)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37546, T)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38189. .38191, AA)
```

```
FT                   /*tag=  at
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38708, A)
FT                   /*tag=  au
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38734. .38736, CG)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38735, A)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41983, G)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42127, T)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42449, A)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42798, G)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(43273, A)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(43743, G)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(45407, C)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(45985, T)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46198, G)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46792. .46194, AG)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46922. .46924, AT)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(47851, G)
FT                   /*tag=  bi
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(48875, T)
FT                   /*tag=  bj
```

Query Match 10.4%; Score 208; DB 10; Length 83450;
Best Local Similarity 77.8%; Pred. No. 1.2e-33;
Matches 277; Conservative 0; Mismatches 75; Indels 4; Gaps 2;

```
Qy      361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
Db    60529 TTATGAATTTCCCTGATGAAAAAGTACAAAATGAGGCCAGGCGTGGTGGCTCATGCCTGT 60588

Qy      421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
Db    60589 AATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGCTCACCTGAGGTCAGGAGTTCAAGACC 60648

Qy      481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
Db    60649 AGCCTGGCCAACATGGTGAAACCTTGTCTCTGCTAAAAATACAAAAATTAGCCAAGCATG 60708

Qy      541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
Db    60709 GTGGCACGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACC 60768

Qy      601 TAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACAG 656
Db    60769 TGGGAGGCAGAGGTTGCATTGAGTCAAGATCACGCCACTGCACTGCAGCCTGGGTGATAG 60828

Qy      657 AGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCT 712
Db    60829 AGCAAGTCTCCATCTCAAAAAAAAAAATTTACAAAGTGGGGCCGGTTGTGGTAGCT 60884
```

```
RESULT 39
ADG88582
ID   ADG88582 standard; DNA; 83450 BP.
XX
AC   ADG88582;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Human phosphatase gene.
XX
KW   Human; phosphatase; immune response; cancer; gene therapy; chromosome 15;
KW   ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(85,T)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1614,G)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(2648,G)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(6773,G)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(8144,G)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9488,C)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9730,G)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10317,C)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10391,C)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10682,A)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(11633,TT)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12054,A)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12938,G)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13719,A)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14332. .14334,TA)
FT                   /*tag=  o
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14753,G)
FT                   /*tag=  p
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14830. .14832,AT)
FT                   /*tag=  q
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
```

```
FT   variation       replace(15157,A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16242,A)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16698,G)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17152,G)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17623. .17625,TA)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17624,T)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18267. .18269,AA)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18786,G)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18812. .18814,CG)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18813,A)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22061,G)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22205,T)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22527,A)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22876,G)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(23351,A)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(23821,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(25487,T)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(30858,C)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(31014,C)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(31437,T)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35341,GA)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35346,G)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36164,A)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36620,G)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37074,G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37545. .37547,TA)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37546,T)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38189. .38191,AA)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38734. .38736,CG)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38735,A)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41983,G)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42127,T)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42449,A)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42798,G)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(43273,A)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(43743,G)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(45407,C)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(45985,T)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46198,G)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46792. .46794,AA)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46922. .46924,AT)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(47851,G)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(48875,T)
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(50844,C)
FT                   /*tag= bi
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(51267,T)
FT                   /*tag= bj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"

  Query Match             10.4%;  Score 208;  DB 10;  Length 83450;
  Best Local Similarity   77.8%;  Pred. No. 1.2e-33;
  Matches  277;  Conservative    0;  Mismatches   75;  Indels    4;  Gaps    2;

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              | ||  | ||   | || ||| ||     ||   |||||||||  || |||| |||||||
Db      60529 TTATGAATTTCCCTGATGAAAAAGTACAAAATGAGGCCAGGCGTGGTGGCTCATGCCTGT 60588

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||||||| ||| || ||||||||||||||||||| ||||||
```

```
Db      60589 AATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGCTCACCTGAGGTCAGGAGTTCAAGACC 60648

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||||||||||||||||| ||||||| |||||||  |||||||||||| |  ||
Db      60649 AGCCTGGCCAACATGGTGAAACCTTGTCTCTGCTAAAAATACAAAAATTAGCCAAGCATG 60708

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              |||||| |||||||||||||| |||||||||||| ||||| ||||||||  ||||||||
Db      60709 GTGGCACGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACC 60768

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACAG 656
              | |||||  ||||||||| | |  | ||||   | ||||||||| |||||| ||  | ||
Db      60769 TGGGAGGCAGAGGTTGCATTGAGTCAAGATCACGCCACTGCACTGCAGCCTGGGTGATAG 60828

Qy        657 AGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCT 712
              |||||| ||||||    | || ||||  ||  ||| || |       || ||| ||
Db      60829 AGCAAGTCTCCATCTCAAAAAAAAAAATTTACAAAGTGGGGCCGGTTGTGGTAGCT 60884


RESULT 40
AEB47870
ID   AEB47870 standard; DNA; 83450 BP.
XX
AC   AEB47870;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Human novel splice form phosphatase genomic DNA.
XX
KW   splice variant; cytostatic; gene therapy; cancer; neoplasm; phosphatase;
KW   ds; gene; chromosome 15; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(85,T)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1614,G)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1698,G)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   CDS             2075. .72181
FT                   /*tag=  d
FT                   /product= "Human novel splice form phosphatase protein"
FT                   /note= "This CDS contains numerous introns, not defined
FT                   within the specification"
FT   variation       replace(2648,G)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(6773,G)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(8144,G)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9488,C)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9730,G)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10317,C)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10391,C)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10682,A)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(11633,-)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12054,A)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12938,G)
FT                   /*tag=  o
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13719,A)
FT                   /*tag=  p
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14333,-)
FT                   /*tag=  q
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14753,G)
FT                   /*tag=  r
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14831,A/-)
FT                   /*tag=  s
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15157,A)
FT                   /*tag=  t
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16242,A)
FT                   /*tag=  u
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17152,G)
FT                   /*tag=  v
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(17624,T/-)
FT                   /*tag=  w
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18268,-)
FT                   /*tag=  x
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18786,G)
FT                   /*tag=  y
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18813,A/-)
FT                   /*tag=  z
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22061,G)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22205,T)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22527,A)
FT                   /*tag=  ac
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(22876,G)
FT                   /*tag=  ad
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(23351,A)
FT                   /*tag=  ae
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(23821,G)
FT                   /*tag=  af
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(25487,T)
FT                   /*tag=  ag
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(30858,C)
FT                   /*tag=  ah
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(31014,C)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(31437,T)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35341,G/A)
```

```
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35346,G)
FT                   /*tag=  al
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36164,A)
FT                   /*tag=  am
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36620,G)
FT                   /*tag=  an
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37074,G)
FT                   /*tag=  ao
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37546,T/-)
FT                   /*tag=  ap
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38190,-)
FT                   /*tag=  aq
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38708,A)
FT                   /*tag=  ar
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38735,A/-)
FT                   /*tag=  as
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41983,A)
FT                   /*tag=  at
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42127,T)
FT                   /*tag=  au
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42449,A)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42798,G)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(43273,A)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(43743,G)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(45407,C)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(45985,T)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46198,G)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46793,-)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46923,-)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(47851,G)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(48875,T)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(50844,C)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(51267,T)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(54073,G)
FT                   /*tag=  bi
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(54206,G)

  Query Match             10.4%;  Score 208;  DB 14;  Length 83450;
  Best Local Similarity   77.8%;  Pred. No. 1.2e-33;
  Matches  277;  Conservative    0;  Mismatches   75;  Indels    4;  Gaps    2;

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              | ||  | ||   | || ||| ||      ||   |||||||| || |||| |||||||
Db      60529 TTATGAATTTCCCTGATGAAAAAGTACAAAATGAGGCCAGGCGTGGTGGCTCATGCCTGT 60588

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||||||| ||| || ||||||||||||||||||||| ||||||
Db      60589 AATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGCTCACCTGAGGTCAGGAGTTCAAGACC 60648

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||||||||||||||||||| ||||||| ||||||  |||||||||||| |  ||
Db      60649 AGCCTGGCCAACATGGTGAAACCTTGTCTCTGCTAAAAATACAAAAATTAGCCAAGCATG 60708

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              |||||| |||||||||||||| |||||||||||| ||||| ||||||||  ||||||||
Db      60709 GTGGCACGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACC 60768

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACAG 656
              | |||||  |||||||| | |  | ||||   | ||||||||| |||||| ||  | ||
Db      60769 TGGGAGGCAGAGGTTGCATTGAGTCAAGATCACGCCACTGCACTGCAGCCTGGGTGATAG 60828

Qy        657 AGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCT 712
              |||||| ||||||    | || ||||  ||  ||| || ||       || ||| ||
Db      60829 AGCAAGTCTCCATCTCAAAAAAAAAAATTTACAAAGTGGGGCCGGTTGTGGTAGCT 60884


RESULT 41
ADQ97960_0/c
WP Sequence split into 4 fragments  LOCUS ADQ97960 Accession Adq97960
WP     Fragment Name      Begin     End
WP     ADQ97960_0             1     110000
WP     ADQ97960_1         100001    210000
WP     ADQ97960_2         200001    310000
WP     ADQ97960_3         300001    390183
ID   ADQ97960 standard; DNA; 390183 BP.
XX
AC   ADQ97960;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer associated sequence HD11-029, SEQ ID 937.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 937; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
```

```
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 390183 BP; 107624A; 72906C; 76848G; 123530T; 0U; 9275Other;

  Query Match             10.4%;  Score 208;  DB 12;  Length 110000;
  Best Local Similarity   72.3%;  Pred. No. 1.3e-33;
  Matches  298;  Conservative    0;  Mismatches  110;  Indels    4;  Gaps    2;

Qy        275 TATGAAATAAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTATTTATTA 334
              | | ||||||  |  |  | |  |  |  | ||| | || |  | | | | || | |
Db      64813 TTTAAAATAAATACATGTATATTTCATTTCCTCAATTTATGAGTGCAAATTAAATTTTAA 64754

Qy        335 CAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATG 394
               || |||   |    | | |  || |||||| |         ||||   |   |   |
Db      64753 GAAATATGCGAACTGAGTATTGAACCTAATATATGACCTAAATTATCAAAAATACCACCT 64694

Qy        395 GGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGA 454
               ||| ||| |||| ||||  |||| |||||||| |||||||||||||||||| |||| ||
Db      64693 AGCCGGGCACAGTGGCTCACGCCTATAATCCCAGCACTTTGGGAGGCCAAGGCGGGCAGA 64634

Qy        455 TCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACT 514
              |||| |||||||||||||| |||||||||||||||||||| |||||||| |||||||||
Db      64633 TCACATGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGCTGAAACCCCGTCTCTACT 64574

Qy        515 AAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAG 574
              |||||  ||||||| ||||||| |||||||    |||||||||||| |||||| ||||
Db      64573 AAAAATACAAAAATCAGCCAGGCGTGGTGGTGCACACCTGTAGTCCCAGCTACTCGGGAG 64514

Qy        575 GTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATG 631
              | | ||| ||||||||  |||||||| ||||| ||||||||||| | || ||   |||
Db      64513 GCTAAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCAAGATCATG 64454

Qy        632 TCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              || |||||||||||||  ||  |||| ||||||||||  |     | || ||||
Db      64453 CCATTGCACTCCAGCCTGGGTGACAGGGCAAGACTCCGTCTCAAAAAAAAAA 64402


RESULT 42
ACN44154/c
ID   ACN44154 standard; DNA; 101209 BP.
XX
AC   ACN44154;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG28567.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 460; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 101209 BP; 21450 A; 25161 C; 26356 G; 23759 T; 0 U; 4483 Other;

  Query Match             10.3%;  Score 207;  DB 11;  Length 101209;
  Best Local Similarity   59.5%;  Pred. No. 2e-33;
  Matches  405;  Conservative    0;  Mismatches  270;  Indels    6;  Gaps    3;

Qy          6 CACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTGGTCGA 65
              ||||  ||| |||||||||||| |||||||||||||| |||||||||| | || || | |
Db      25000 CACCCACCTCGGCCTCCCAAAGCGCTGGGATTACAGGTGTGAGCCACCGCACCCGGCCCA 24941

Qy         66 ATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGACAGACT 125
                |  |  |           |   |   |  |  ||          | |  |  |
Db      24940 GATTGGCTTCTTCCACGTAGCAATGTACATTTCAGCATCCTCCACATCTGCTCAAGGCTT 24881

Qy        126 TTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAGTCTCT 185
              |     |   |  |  |  |     || |     ||   ||   |     ||
Db      24880 GACAGCTCATTTCCTTTTATCCATAAACAATATTCCCTTGCAGGGAAGCGCCACCGTTCG 24821

Qy        186 CTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAATAAAC 245
              | || |   |   |  || | |  |    || |  |      |   |  | | |||| ||
Db      24820 CACATCCACCCACCACCACAGGGTATCCTGACTGCTTCCACCGCGCCCGTATGAATGAAG 24761

Qy        246 CTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAGTAGCT 305
              | | ||         |     |||| || |||  | || ||| |||| |   |
Db      24760 TGGTTACAAACACCTGCGTGCAGGTTTCATGTGAAGGATGGTTTTGAACATTTTTAAATG 24701

Qy        306 GCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATA 365
              |   ||      ||   || | | |||||   | || |  || | |||  | | |
Db      24700 TTGAAAAAAATCCAGAGGAGAATATTTCACAACACATGAAAATTATATGAAATTCCACTT 24641

Qy        366 CACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCC 425
              |  ||  |||| |    ||||  ||||  ||||| || ||||  |||||||| ||
Db      24640 TCAGT--ATCCTTAATAAAACGTGAATGGGGCTGGGCGCGGTGGCTCACGCCTGTAACCC 24583

Qy        426 CAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCT 485
              || |||| ||||||||| ||| ||| |||| |||||||||||||| ||| |||||||||
Db      24582 CAGCACTCTGGGAGGCCGAGGCGGGGGGATTACCTGAGGTCAGGAATTTGAGACCAGCCC 24523

Qy        486 GGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGC 545
              |||||||| || |||||||  |||||| ||||||  |||||||||||  |||||||||
Db      24522 GGCCAACACGGCGAAACCCCATCTCTAGTAAAAATACAAAAATTAGCCGTGTGTGGTGGT 24463

Qy        546 ATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGA 605
              |||||||||||||||| |||||| ||||| ||||| ||||||||||||  |||  ||||
Db      24462 GGGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTTAACTCAGGA 24403

Qy        606 GGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACAGAGCAA 661
              |||||||||||||| | || ||||     |||||||||||||||| ||| | ||||  |
Db      24402 GGTGGAGGTTGCAGTGAGCCAAGATCACACCACTGCACTCCAGCCTGGGCGAGAGAGTGA 24343

Qy        662 GACTCCATAAAGACAACAAAA 682
              ||| |  |    | || ||||
Db      24342 GACACTGTCTCAAAAAAAAAA 24322


RESULT 43
ABD33366
```

```
ID   ABD33366 standard; DNA; 243390 BP.
XX
AC   ABD33366;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-067.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 462; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 243390 BP; 65851 A; 48827 C; 53197 G; 74003 T; 0 U; 1512 Other;

  Query Match             10.3%;  Score 207;  DB 13;  Length 243390;
  Best Local Similarity   79.2%;  Pred. No. 2.3e-33;
  Matches  259;  Conservative    0;  Mismatches   65;  Indels    3;  Gaps    1;

Qy        359 ACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCT 418
              | || |   |  | ||| || || |      || || ||| ||||| |  ||||   |||
Db      74590 AGTATTTGGCAGACATTTTTGAAAATGAACAAAGTGAGCCGGGCGCGGGGGCTCACACCT 74649

Qy        419 GTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGA 478
              ||||||||| |||||||||||||||||||||||| |||||||||||||| |||||| ||||
Db      74650 GTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCCGATCACCTGAGGTCCGGAGTTCAAGA 74709

Qy        479 CCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTG 538
              |||||||||||||||||||||||| ||||||||||||||  ||||||||||  ||||
Db      74710 CCAGCCTGGCCAACATGGTGAAACCCAGTCTCTACTAAAAATACAAAAATTAGCTGGGTG 74769

Qy        539 TGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAA 598
              |||||||||||||||||||| || |||||||||||| ||||| ||||||||||||||||
Db      74770 TGGTGGCATGCACCTGTAGTCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAA 74829

Qy        599 CCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGCAACA 655
              ||  |||||||||||||||||| | | ||||   || || ||||||||||||||  |
Db      74830 CCCTGGAGGTGGAGGTTGCAGTGAACTGAGATGGTGCCATTGCACTCCAGCCTGGGCAAC 74889

Qy        656 GAGCAAGACTCCATAAAGACAACAAAA 682
                ||   ||  |  |     || ||||
Db      74890 AAGAGTGAAACTCCATTTCAAAAAAAA 74916


RESULT 44
AAK70746/c
ID   AAK70746 standard; DNA; 22700 BP.
XX
AC   AAK70746;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25558.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
```

PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 25558; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 22700 BP; 5667 A; 5046 C; 5162 G; 6825 T; 0 U; 0 Other;

Query Match 10.3%; Score 206.8; DB 4; Length 22700;
Best Local Similarity 82.0%; Pred. No. 1.8e-33;
Matches 251; Conservative 0; Mismatches 52; Indels 3; Gaps 1;

Qy 380 AAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAG 439
Db 3358 AAGAGCAAAACTCTGAGCCAGGCGTGGTGCCTCATGACTGTAATCCCAGTCCTTTGGGAG 3299

Qy 440 GCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGA 499
Db 3298 GCCGAGGAGGGCAGATCACCCGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGA 3239

Qy 500 AACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTC 559
Db 3238 AACCCTGTCTCTACTAAAAATACAAAAATTAGCCGGATGTGGTGGCACATTCCAGTAATC 3179

Qy 560 CCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAG 619
Db 3178 TCAGCTACTCAGGAAGCTGAGGCAGGAGAATTGCTAGAACCTGGGAGGCAGAGGTTACAG 3119

Qy 620 TAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACA 676
Db 3118 TGAGCCGAGATCACACCACTGCACTCCAGCCTGGCAACAGAGCAAAACTCCATCTAAAAA 3059

```
Qy        677 ACAAAA 682
                | ||||
Db       3058 AAAAAA 3053


RESULT 45
ABD32596
ID   ABD32596 standard; DNA; 59365 BP.
XX
AC   ABD32596;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD7-220.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 100; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 59365 BP; 19132 A; 10694 C; 10458 G; 19061 T; 0 U; 20 Other;

  Query Match             10.3%;  Score 206.8;  DB 13;  Length 59365;
  Best Local Similarity   74.9%;  Pred. No. 2.1e-33;
  Matches  286;  Conservative    0;  Mismatches   92;  Indels    4;  Gaps    2;

Qy        376 ATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTG 435
              || ||| |   |  || |  ||  || || || ||||  |||||||||||| |||||||
Db      46577 ATCAAAAATGGTGAAATTCTGCTGGGTGCGGTGGCTCACGCCTGTAATCCCAGCACTTTG 46636

Qy        436 GGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATG 495
              ||||||  ||| ||| ||||| |||||||||||||| | ||||||||||||||||||||
Db      46637 GGAGGCTGAGGAGGGTGGATCGCCTGAGGTCAGGAGCTCAAGACCAGCCTGGCCAACATG 46696

Qy        496 GTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGT 555
              |||||||||  ||||||||||||  ||||||||||  |||||||| | ||||||| |
Db      46697 GTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGACGTGCACCTAT 46756

Qy        556 AGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTT 615
              | ||||| |||||| |||||||||| |||||||| |||||| ||| ||||||||||||
Db      46757 AATCCCAGCTACTCGGGAGGTTGAGGCAGGAGAATGGCTTGAGCCTGGGAGGTGGAGGTT 46816

Qy        616 GCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGG-CAACAGAGCAAGACTCCATAA 671
              ||||| | || ||||   | ||||||||||||||||| |||| ||||||||||   |
Db      46817 GCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGACAACTGAGCAAGACTGTGTCC 46876

Qy        672 AGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGAAATTC 731
                | || ||||      |||      ||| ||  | | | | || | | |      |||
Db      46877 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAAGTCTGCATTTCATTACTCTTTTCATTG 46936

Qy        732 ATCTTTGTTCATTTATTTTTAC 753
              || ||| | |   | |||| ||
Db      46937 ATTTTTCTCCTACTTTTTTCAC 46958


Search completed: January 22, 2006, 04:08:52
Job time : 819.081 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:39:03 ; Search time 245.225 Seconds
(without alignments)
14504.671 Million cell updates/sec

Title: US-09-728-552A-3_COPY_39000_41000
Perfect score: 2001
Sequence: 1 tgatccaccagccttggcct.....ctaattcagtgaacacttca 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	80595	3	US-09-078-294-3	Sequence 3, Appli
	2	1959.4	97.9	80246	3	US-09-078-294-4	Sequence 4, Appli
c	3	223.8	11.2	393753	3	US-09-949-016-14573	Sequence 14573, A
c	4	223.8	11.2	393753	3	US-09-949-016-14574	Sequence 14574, A
c	5	223.8	11.2	818128	3	US-09-949-016-14546	Sequence 14546, A
c	6	223.8	11.2	818128	3	US-09-949-016-14547	Sequence 14547, A
c	7	223.8	11.2	818128	3	US-09-949-016-14548	Sequence 14548, A
c	8	223.8	11.2	818128	3	US-09-949-016-14549	Sequence 14549, A
c	9	223.8	11.2	818128	3	US-09-949-016-14550	Sequence 14550, A
c	10	223.8	11.2	818128	3	US-09-949-016-14551	Sequence 14551, A
c	11	223.8	11.2	818128	3	US-09-949-016-14552	Sequence 14552, A
c	12	223.8	11.2	818128	3	US-09-949-016-14553	Sequence 14553, A
c	13	223.8	11.2	818128	3	US-09-949-016-14554	Sequence 14554, A
c	14	223.8	11.2	818128	3	US-09-949-016-14555	Sequence 14555, A
c	15	223.8	11.2	818128	3	US-09-949-016-14556	Sequence 14556, A
c	16	223.8	11.2	818128	3	US-09-949-016-14557	Sequence 14557, A
c	17	223.8	11.2	818128	3	US-09-949-016-14558	Sequence 14558, A
c	18	223.8	11.2	818128	3	US-09-949-016-14559	Sequence 14559, A
c	19	223.8	11.2	818128	3	US-09-949-016-14560	Sequence 14560, A
c	20	223.8	11.2	818128	3	US-09-949-016-14561	Sequence 14561, A
c	21	223.8	11.2	818128	3	US-09-949-016-14562	Sequence 14562, A
c	22	223.8	11.2	818128	3	US-09-949-016-14564	Sequence 14564, A
c	23	223.8	11.2	818128	3	US-09-949-016-14565	Sequence 14565, A
c	24	223.8	11.2	818128	3	US-09-949-016-14566	Sequence 14566, A
c	25	223.8	11.2	818128	3	US-09-949-016-14567	Sequence 14567, A
c	26	217.6	10.9	64813	3	US-09-949-016-11957	Sequence 11957, A
c	27	217.6	10.9	70131	3	US-09-949-016-16064	Sequence 16064, A
c	28	214.4	10.7	325791	3	US-09-768-185A-1	Sequence 1, Appli
	29	212.2	10.6	43726	3	US-09-949-016-17578	Sequence 17578, A
	30	212.2	10.6	76962	3	US-09-949-016-17482	Sequence 17482, A
	31	211.8	10.6	45842	3	US-09-949-016-12550	Sequence 12550, A
	32	211.8	10.6	45842	3	US-09-949-016-17327	Sequence 17327, A
c	33	211.6	10.6	264206	3	US-09-949-016-12731	Sequence 12731, A
c	34	211.6	10.6	264304	3	US-09-949-016-13249	Sequence 13249, A
	35	211.4	10.6	54286	3	US-09-949-002-837	Sequence 837, App
	36	211.4	10.6	66955	3	US-09-949-002-642	Sequence 642, App
c	37	210.8	10.5	12579	3	US-09-949-016-17270	Sequence 17270, A
c	38	210.6	10.5	63187	3	US-09-949-016-12682	Sequence 12682, A
c	39	210.6	10.5	63187	3	US-09-949-016-16288	Sequence 16288, A
c	40	210.6	10.5	177669	3	US-09-949-016-13713	Sequence 13713, A
	41	210.4	10.5	42053	3	US-09-949-016-15924	Sequence 15924, A
c	42	210.2	10.5	283538	3	US-09-949-016-13506	Sequence 13506, A
c	43	210	10.5	78630	3	US-09-949-016-16790	Sequence 16790, A
c	44	209.4	10.5	78846	3	US-09-949-016-12396	Sequence 12396, A
c	45	209.4	10.5	78846	3	US-09-949-016-12791	Sequence 12791, A
c	46	209.4	10.5	78846	3	US-09-949-016-12792	Sequence 12792, A
c	47	209.4	10.5	78846	3	US-09-949-016-12793	Sequence 12793, A
c	48	209.4	10.5	78850	3	US-09-949-016-16013	Sequence 16013, A
c	49	209.4	10.5	78850	3	US-09-949-016-16014	Sequence 16014, A
c	50	209.4	10.5	78850	3	US-09-949-016-16015	Sequence 16015, A
c	51	209.4	10.5	78850	3	US-09-949-016-16016	Sequence 16016, A
c	52	209.4	10.5	78850	3	US-09-949-016-16201	Sequence 16201, A
c	53	209.4	10.5	78850	3	US-09-949-016-16202	Sequence 16202, A
c	54	209.4	10.5	78850	3	US-09-949-016-16203	Sequence 16203, A
c	55	209.4	10.5	78850	3	US-09-949-016-16204	Sequence 16204, A
	56	209.2	10.5	144596	3	US-09-949-016-11749	Sequence 11749, A
	57	209.2	10.5	144596	3	US-09-949-016-13035	Sequence 13035, A
	58	208	10.4	83450	3	US-09-811-469-3	Sequence 3, Appli
	59	208	10.4	83450	3	US-10-370-659-3	Sequence 3, Appli
c	60	207.6	10.4	61108	3	US-09-949-002-608	Sequence 608, App
	61	207.6	10.4	64984	3	US-09-949-016-15254	Sequence 15254, A
	62	207.6	10.4	205044	3	US-09-949-016-15851	Sequence 15851, A
	63	207.6	10.4	205044	3	US-09-949-016-15852	Sequence 15852, A
	64	207.6	10.4	205044	3	US-09-949-016-15853	Sequence 15853, A
	65	207.6	10.4	223471	3	US-09-949-016-12387	Sequence 12387, A
	66	207.6	10.4	223471	3	US-09-949-016-12724	Sequence 12724, A
	67	207.6	10.4	223471	3	US-09-949-016-12725	Sequence 12725, A
c	68	207.4	10.4	601	3	US-09-949-016-21480	Sequence 21480, A
c	69	207.4	10.4	601	3	US-09-949-016-91386	Sequence 91386, A
c	70	207.4	10.4	41612	3	US-09-949-016-12769	Sequence 12769, A
c	71	207.4	10.4	41639	3	US-09-949-016-12773	Sequence 12773, A
	72	207	10.3	63783	3	US-09-949-016-13576	Sequence 13576, A
	73	207	10.3	227390	3	US-09-949-016-12201	Sequence 12201, A
	74	207	10.3	227391	3	US-09-949-016-13365	Sequence 13365, A
	75	206.8	10.3	43366	3	US-09-949-016-14185	Sequence 14185, A
	76	206.8	10.3	107458	3	US-09-949-016-15687	Sequence 15687, A
	77	206.6	10.3	13533	3	US-09-949-016-17162	Sequence 17162, A
	78	206.6	10.3	125902	3	US-09-949-016-13715	Sequence 13715, A
	79	206.4	10.3	27923	3	US-09-949-016-13371	Sequence 13371, A
	80	206.2	10.3	601	3	US-09-949-016-187035	Sequence 187035,
	81	206.2	10.3	41863	3	US-09-949-016-14948	Sequence 14948, A
c	82	206.2	10.3	81001	3	US-09-750-580-1	Sequence 1, Appli
	83	206.2	10.3	117001	3	US-09-949-016-15684	Sequence 15684, A
c	84	206.2	10.3	143155	3	US-09-949-016-11925	Sequence 11925, A
c	85	206.2	10.3	143164	3	US-09-949-016-14368	Sequence 14368, A
c	86	206.2	10.3	143173	3	US-09-949-016-14513	Sequence 14513, A
	87	206	10.3	183202	3	US-09-949-016-13614	Sequence 13614, A
c	88	206	10.3	636591	3	US-09-949-016-11808	Sequence 11808, A
c	89	206	10.3	636591	3	US-09-949-016-13388	Sequence 13388, A
	90	205.8	10.3	27465	3	US-09-949-016-16561	Sequence 16561, A
c	91	205.8	10.3	57507	3	US-09-949-016-15019	Sequence 15019, A
c	92	205.8	10.3	246444	3	US-09-949-016-13113	Sequence 13113, A
	93	205.6	10.3	601	3	US-09-949-016-107997	Sequence 107997,
c	94	205.6	10.3	53769	3	US-09-949-016-17527	Sequence 17527, A
c	95	205.6	10.3	93398	3	US-09-949-016-14167	Sequence 14167, A
c	96	205.4	10.3	601	3	US-09-949-016-118216	Sequence 118216,
	97	205.4	10.3	63385	3	US-09-949-016-14856	Sequence 14856, A

c 98	205.4	10.3	90618	3	US-09-949-016-15964	Sequence 15964, A
c 99	205.2	10.3	91933	3	US-09-949-016-11855	Sequence 11855, A
c 100	205.2	10.3	91933	3	US-09-949-016-14628	Sequence 14628, A

ALIGNMENTS

RESULT 1
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match 100.0%; Score 2001; DB 3; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39000 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 39059

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39060 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 39119

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39120 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 39179

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39180 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 39239

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39240 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 39299

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39300 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39359

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39360 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39419

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39420 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39479

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39480 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39539

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39540 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39599

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39600 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39659

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39660 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39719

Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39720 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39779

Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39780 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39839

Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39840 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39899

Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39900 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 39959

Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39960 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 40019

Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40020 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 40079

Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40080 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 40139

Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40140 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 40199

Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40200 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 40259

Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40260 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 40319

Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40320 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40379

Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40380 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40439

Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40440 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40499

Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40500 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 40559

Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40560 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 40619

Qy       1621 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40620 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 40679

Qy       1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      40680 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40739

Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40740 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40799

Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40800 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40859

Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40860 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 40919

Qy       1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40920 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 40979

Qy       1981 TCTAATTCAGTGAACACTTCA 2001
              |||||||||||||||||||||
Db      40980 TCTAATTCAGTGAACACTTCA 41000


RESULT 2
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

  Query Match             97.9%;  Score 1959.4;  DB 3;  Length 80246;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1988;  Conservative    0;  Mismatches   11;  Indels    2;  Gaps    2;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38738 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 38797

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38798 GTCGAATGTTTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 38857

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38858 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 38917

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db      38918 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGC-TTCAAGGAGGAA 38976

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||
Db      38977 TAAACCTATGAAAAGTGTTGAGGATAGTGTCTGATATGAAATAAGGA-TTCAACAAGTAG 39035

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39036 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 39095

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      39096 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 39155

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39156 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 39215

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39216 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 39275

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39276 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 39335

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              |||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      39336 TAGGAGCTGGAGGTTGCAGTGACCCGAGATGTCACTGCACTCCAGCCTGGCAACAGAGCA 39395

Qy        661 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39396 AGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCAT 39455

Qy        721 TTAAGAAATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 780
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      39456 TTAAGAAATTCATCTTTGTTCATCTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTT 39515

Qy        781 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39516 AATTAAGCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATTATTTGCTAAAA 39575

Qy        841 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39576 TAATATTCTTGCTTCTTTTTTAAGGAATTGTCTCCCTAGAAACGGTTTGTACCAAACAAT 39635

Qy        901 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAAGAT 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      39636 ACACTGACTTTACACAAAATCAGATCTGATTGGCAACAGTTGCAGATGTTTTCAAAGGAT 39695

Qy        961 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39696 TTTCATTTGAGAAGGGGCCCATTTGGGTTATTTAGATTCTAAGAACTGAAACTGCTTTGT 39755

Qy       1021 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39756 TCTGTTTTTCTGGCTTCTGGGAGAGGAGGAGACATGAATTCAGTTAGCACCTTGGTATTT 39815

Qy       1081 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39816 TCTTTATCCTTCATTTCAATACAGAAGATGCTTCATATGCACAGTGGTGTCAGGTCACAT 39875

Qy       1141 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39876 CAAAAGAAAGAGAAACAGTTTCTTGGTTTTTAATTTTCAACCGGAAAGGAAAGGCACCCA 39935

Qy       1201 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39936 TTTTGTTCCGCTCTAATTAGCCAGTGCATGACTTAGAGAGCAGGCAGATGCTTTGAAGGC 39995

Qy       1261 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39996 GTGGTAACACAGGTCTTCATTAATCTCCACGCAGGACTTGCACTTCTACTATGCCTAGGC 40055

Qy       1321 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40056 TGAAGAAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAACTAACTGAAGCCAGG 40115

Qy       1381 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40116 TGTTATAAAGCACAAGTCAAGAGGGTGAGAAACTAACGTTCTTGAAATCTCCCACTTCTT 40175

Qy       1441 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40176 TCTACGTCAGAAGAGCCAAGCTGATTATTTTAGTTGGAATTTAGAAATTTTTAAAAATTA 40235
```

```
Qy       1501 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40236 TTCTAAAGTCATGAACAAGCCTAATTATAAAGATAGTTGCTGTGAAGGTGCTGAAATAAC 40295

Qy       1561 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCATAATGGAATCCTGTACAATGTTCAC 1620
             |||||||||||||||||||||||||||||||||| |||||||||||||| |||||||||
Db      40296 TCGATTTTACCAACCCCCTCTTCTGGAGGAAGCCAGAATGGAATCCTGTAGAATGTTCAC 40355

Qy       1621 TCTACCAACGAACTCTTGTTTTTTCTAATGAGGAAACAGAGGCCCACAGTATTAAACTATC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      40356 TCTACCAACGAACTCTTGTTTTTCTAATGAGGAAACAGAGGCCCACAGTAGTAAACTATC 40415

Qy       1681 TTAACCAATACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 1740
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40416 TTAACCAAGACAAAATGACTAGTGCTCTGGTCCTTTTATTAAGCACTAAAATTTTGATCC 40475

Qy       1741 AATAATAAATCTGTCCATTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 1800
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      40476 AATAATAAATCTGTCCAGTAGAAGGAGTTTCCCTAATGTACTGGTTCTAACTTGTTCCCT 40535

Qy       1801 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40536 TCAAGGGGCCAGTGTCCCGTACACATAGCTAAATGGGACTTCTCTTCAACTACCATTACC 40595

Qy       1861 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40596 CAGAGGGCAGAACCTAAAATGCTGTGAATGACATTCTGCTGTTCACATCTCAGCAGCAGT 40655

Qy       1921 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40656 GTTGCATTTGAGCTTCTGCAGGGCCACCCAGGACCTATATCTGCTCAGATGTTTAACTCA 40715

Qy       1981 TCTAATTCAGTGAACACTTCA 2001
             |||||||||||||||||||||
Db      40716 TCTAATTCAGTGAACACTTCA 40736
```

RESULT 3
US-09-949-016-14573/c
; Sequence 14573, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14573
; LENGTH: 393753
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(393753)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14573

Query Match 11.2%; Score 223.8; DB 3; Length 393753;
Best Local Similarity 61.3%; Pred. No. 3.4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||| |
Db     122074 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 122015

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||     |       ||| |  | |  | | |
Db     122014 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 121955

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |     ||| || |   || | | |     |   |       | |
Db     121954 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 121895

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| ||| |   |     |  |  |||| ||
Db     121894 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 121835

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||                ||  |   | || |   | |  || |   || |   | || ||
Db     121834 TTAG-------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 121788

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | |  | |||  | | |  |  |     |
Db     121787 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 121730

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              |   |       | | || |   || ||   ||||||| || || |||| ||||||
Db     121729 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 121670

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db     121669 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 121610

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||  ||| || |||||||||||||  |||||||||||||||  |
Db     121609 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 121550

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| |||||  || ||  ||||| ||||| ||||||||||||||||
Db     121549 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 121490

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||||| | || ||||    | ||||||||||||||| ||  |||
Db     121489 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 121430

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  |||||||  |  || |||  |     ||     |||  | |||   | ||
Db     121429 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 121370

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     121369 GGAAGATCTGAATCTCAGCAT 121349
```

RESULT 4
US-09-949-016-14574/c
; Sequence 14574, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14574
; LENGTH: 393753
; TYPE: DNA

```
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(393753)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14574

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 393753;
  Best Local Similarity   61.3%;  Pred. No. 3.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| |||||||||||||| |
Db     122074 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 122015

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||     |       ||| |   | |   | | |
Db     122014 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 121955

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |       ||| || |   || | | |     |   |       | |
Db     121954 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 121895

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |   || ||  || |||| |||  |   |     |  |  |||| ||
Db     121894 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 121835

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||                ||  |   | || |   | |  || |   || |   | || ||
Db     121834 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 121788

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||   || ||  ||| | |  | |||  | |  |  |   |     |
Db     121787 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 121730

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              |    |         | | || |   || ||   ||||||| || || |||| ||||||
Db     121729 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 121670

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db     121669 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 121610

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              |||||| |||||||||  ||| || ||||||||||||||  ||||||||||||||| | |
Db     121609 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 121550

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db     121549 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 121490

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||| | || ||||   | ||||||||||||||| ||  |||
Db     121489 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 121430

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |      ||       ||| | |||   | ||
Db     121429 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 121370

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     121369 GGAAGATCTGAATCTCAGCAT 121349


RESULT 5
US-09-949-016-14546/c
; Sequence 14546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14546
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14546

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| |||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||     |       ||| |   | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |       ||| || |   || | | |     |   |       | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |   || ||  || |||| |||  |   |     |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||                ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||   || ||  ||| | |  | |||  | |  |  |   |     |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              |    |         | | || |   || ||   ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              |||||||  |||||||||  ||| || ||||||||||||||  |||||||||||||| | |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||| | || ||||   | ||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |      ||       ||| | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
```

```
                  |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724
```

RESULT 6
US-09-949-016-14547/c
; Sequence 14547, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14547
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14547

  Query Match   11.2%; Score 223.8; DB 3; Length 818128;
  Best Local Similarity 61.3%; Pred. No. 4.4e-45;
  Matches 454; Conservative   0; Mismatches 267; Indels   20; Gaps   5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |      ||| |   | |   | |  |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |        ||| || |   || | | |      |   |        | ||
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |  |   || ||   || |||| |||  |    |    |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||              ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | |   | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              |  |   |  | || |   || ||   ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| ||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||  ||| || |||||||||||||  |||||||||||||| |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925
```

```
Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |  |||||||||||||||| | || ||||   | ||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  |||||||  |    | ||  ||| |     ||       |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724
```

RESULT 7
US-09-949-016-14548/c
; Sequence 14548, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14548
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14548

  Query Match   11.2%; Score 223.8; DB 3; Length 818128;
  Best Local Similarity 61.3%; Pred. No. 4.4e-45;
  Matches 454; Conservative   0; Mismatches 267; Indels   20; Gaps   5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |      ||| |   | |   | |  |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |        ||| || |   || | | |      |   |        | ||
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |  |   || ||   || |||| |||  |    |    |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||              ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | |   | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105
```

```
Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
Db     545744 GGAAGATCTGAATCTCAGCAT 545724
```

RESULT 8
US-09-949-016-14549/c
; Sequence 14549, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14549
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14549

Query Match 11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4.4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
```

```
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
Db     545744 GGAAGATCTGAATCTCAGCAT 545724
```

RESULT 9
US-09-949-016-14550/c
; Sequence 14550, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14550
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14550

Query Match 11.2%; Score 223.8; DB 3; Length 818128;

```
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||||| |||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||     |       ||| |   | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |  |    ||| || |   || | | |     |   |       | | |  | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||  || |||| |||  |   |       | |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||                ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | |  | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
                |    |         | | || |   || ||  ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||  ||| || ||||||||||||||  ||||||||||||||  |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| |||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||| | || ||||    | ||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |     ||      |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 10
US-09-949-016-14551/c
; Sequence 14551, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14551
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14551

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||||| |||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||     |       ||| |   | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |  |    ||| || |   || | | |     |   |       | | |  | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||  || |||| |||  |   |       | |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||                ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | |  | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
                |    |         | | || |   || ||  ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||  ||| || ||||||||||||||  ||||||||||||||  |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| |||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||| | || ||||    | ||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |     ||      |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 11
US-09-949-016-14552/c
; Sequence 14552, Application US/09949016
; Patent No. 6812339
```

```
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14552
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14552

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||     |       ||| |   | |   | | | :
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                  |    |     ||| || |   || | | |     |   |         | ||
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| |||  |   |     | |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||                |  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | | | |||  | |  |   |  |     |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              |   |     | | || |   || ||   ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||  ||| || ||||||||||||||  |||||||||||||||  |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |  ||||||||||||||||| | || ||||   | ||||||||||||||| || |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  |||||||  |    | ||  ||| |      ||      |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  | |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 12
US-09-949-016-14553/c
; Sequence 14553, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14553
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14553

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||     |       ||| |   | |   | | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                  |    |     ||| || |   || | | |     |   |         | ||
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| |||  |   |     | |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||                |  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | | | |||  | |  |   |  |     |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              |   |     | | || |   || ||   ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985
```

```
Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||  ||| || ||||||||||||||  ||||||||||||||  |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| |||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||| | || ||||   | |||||||||||||||  ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |     ||      |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 13
US-09-949-016-14554/c
; Sequence 14554, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14554
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14554

  Query Match            11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || ||| ||||     |       ||| |   | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |       ||| || |   || | | |       |   |         | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| ||| |   |     |   |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
                ||              ||  |   | || |   | |  || |   || |   | || ||
```

```
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||     || ||  ||| | |  | ||| | |  |  |  |  |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              |   |       | | || |   || ||   ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| ||||||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||  ||| || ||||||||||||||  |||||||||||||||  |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| |||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||| | || ||||   | |||||||||||||||  ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |     ||      |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 14
US-09-949-016-14555/c
; Sequence 14555, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14555
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14555

  Query Match            11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
```

```
Db      546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy         120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
               |   |        ||| || |   || | |    |   |        | |
Db      546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy         180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
               | |    |    || ||   || |||| ||| |    |      |  |  |||| ||
Db      546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy         240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
               ||                 ||  |   | || |   | |  || |   || |   | || ||
Db      546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy         300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
               |||  |   |||  |||    || ||  ||| | |   | |||   | |  |  |    |
Db      546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy         360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
               |    |         | | ||  |   || ||   ||||||| || || |||| ||||||
Db      546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy         420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
               |||||||| |||||| ||||||||||||||||||||||||||||||||||||||| ||||
Db      546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy         480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
               ||||||| |||||||||   ||| || ||||||||||||||  |||||||||||||   |
Db      545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy         540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
               ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db      545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy         600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
               |   |||||||||||||||| | || ||||   | ||||||||||||||| ||  |||
Db      545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy         656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
               |||  |||||||  |    | ||  |||  |      ||       ||| | |||   | ||
Db      545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy         716 TTCATTTAAGAAATTCATCTT 736
               |  |  |||  ||| | |
Db      545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 15
US-09-949-016-14556/c
; Sequence 14556, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14556
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14556

  Query Match           11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity  61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy           1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
               ||||||||| |||||||||||||||||||||| |||||||| |||||||||||||| |
Db      546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy          61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
               | | |   ||| | || ||| ||||     |      ||| |   | |   | |  |
Db      546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy         120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
               |   |      ||| || |  || | |   |      |  |  | | |
Db      546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy         180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
               | |    |   || ||   || |||| |||  |    |      |  |  |||| ||
Db      546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy         240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
               ||                 ||  |   | || |   | |  || |   || |   | || ||
Db      546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy         300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
               |||  |   |||  |||    || ||  ||| | |   | |||   | |  |  |    |
Db      546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy         360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
               |    |         | | ||  |   || ||   ||||||| || || |||| ||||||
Db      546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy         420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
               |||||||| |||||| ||||||||||||||||||||||||||||||||||||||| ||||
Db      546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy         480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
               |||||||  |||||||||   ||| || ||||||||||||||  |||||||||||||   |
Db      545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy         540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
               ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db      545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy         600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
               |   |||||||||||||||| | || ||||   | ||||||||||||||| ||  |||
Db      545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy         656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
               |||  |||||||  |    | ||  |||  |      ||       ||| | |||   | ||
Db      545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy         716 TTCATTTAAGAAATTCATCTT 736
               |  |  |||  ||| | |
Db      545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 16
US-09-949-016-14557/c
; Sequence 14557, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
```

```
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14557
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14557

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||||| ||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |     ||| |  | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |       ||| || |   || | | |     |   |        | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| |||  |   |      |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||                ||  |  | || |  | |  || |  || |  | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | |  | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
                |    |       | | || |   || ||   |||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||   ||| || ||||||||||||||   |||||||||||||   |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |  ||||||||||||||||| | || ||||   | ||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |      ||      |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  | |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724
```

```
RESULT 17
US-09-949-016-14558/c
; Sequence 14558, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14558
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14558

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||||| ||||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |     ||| |  | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |       ||| || |   || | | |     |   |        | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| |||  |   |      |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||                ||  |  | || |  | |  || |  || |  | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | |  | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
                |    |       | | || |   || ||   |||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||   ||| || ||||||||||||||   |||||||||||||   |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865
```

```
Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||||| | || ||||    | ||||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |     ||       |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 18
US-09-949-016-14559/c
; Sequence 14559, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14559
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14559

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||||||| |||||||||||||||||||||| |||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |       ||| |   | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |       ||| || |   || | | |     |   |         | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| |||  |   |     |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||          ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              |||  |   |||  |||    || ||  ||| | |  | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
                |    |         | | || |   || ||   ||||||| || || |||| ||||||

Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| ||||||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||   ||| || ||||||||||||||  ||||||||||||||  |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||||| | || ||||    | ||||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |     ||       |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 19
US-09-949-016-14560/c
; Sequence 14560, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14560
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14560

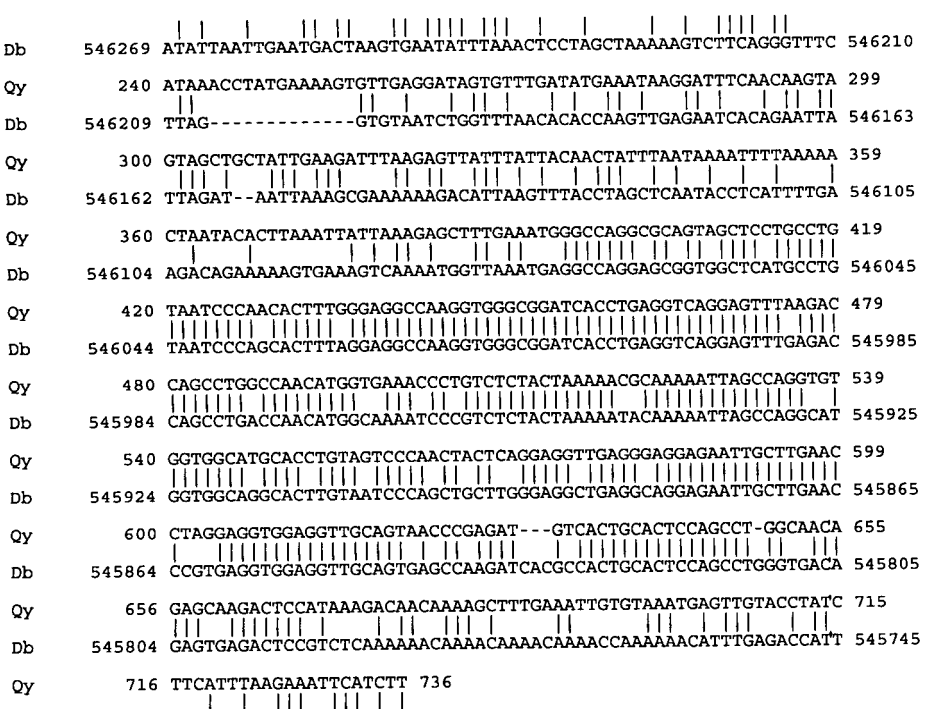
  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||||||| |||||||||||||||||||||| |||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |       ||| |   | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                |   |       ||| || |   || | | |     |   |         | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
```



```
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299

Db     546209 TTAG-------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359

Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419

Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479

Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539

Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599

Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655

Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715

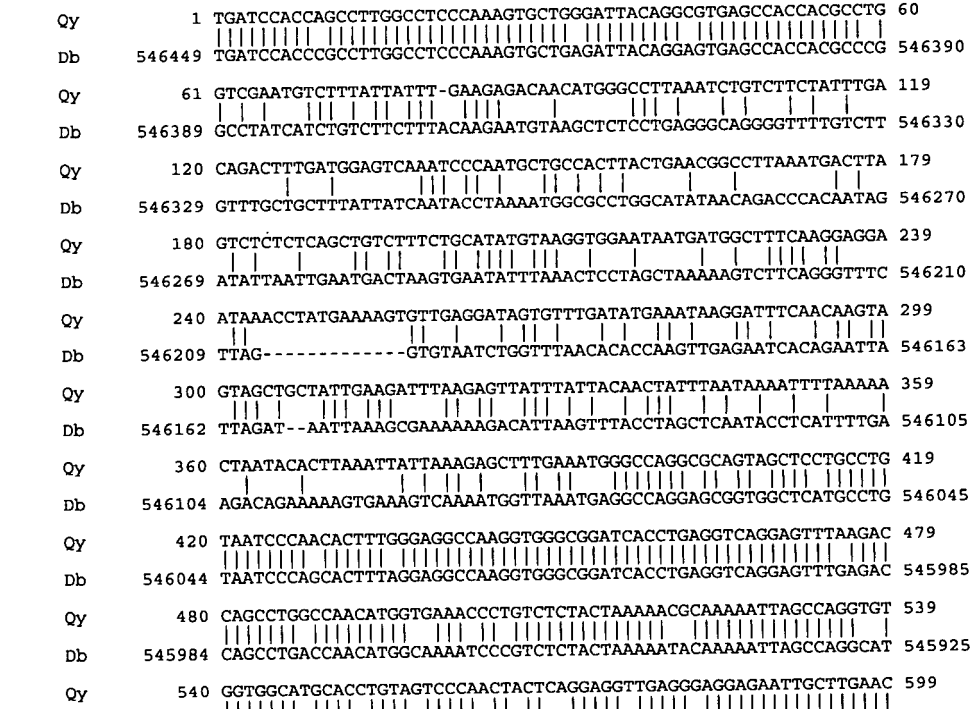
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736

Db     545744 GGAAGATCTGAATCTCAGCAT 545724
```

```
RESULT 20
US-09-949-016-14561/c
; Sequence 14561, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14561
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14561

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;
```



```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60

Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119

Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179

Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239

Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299

Db     546209 TTAG-------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359

Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419

Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479

Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539

Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599

Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655

Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715

Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736

Db     545744 GGAAGATCTGAATCTCAGCAT 545724
```

```
RESULT 21
US-09-949-016-14562/c
; Sequence 14562, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14562
```

```
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14562

  Query Match            11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| |||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |       ||| |   | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                  |   |       ||| || |   || | | |       |   |         | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
                | |   |   || ||   || |||| |||  |   |     |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
                ||              ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
                ||| |   ||| |||    || ||  ||| | |  | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
                |   |       | | || |   || ||   ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
                |||||||| |||||| ||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
                ||||||| |||||||||   ||| || ||||||||||||||   ||||||||||||||  |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
                ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
                |   |||||||||||||||| | || ||||    | ||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
                |||  ||||||| |    | ||  ||| |     ||       |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
                |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 22
US-09-949-016-14564/c
; Sequence 14564, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14564
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14564

  Query Match            11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| |||||||| ||||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |       ||| |   | |   | | |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                  |   |       ||| || |   || | | |       |   |         | |
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
                | |   |   || ||   || |||| |||  |   |     |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
                ||              ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
                ||| |   ||| |||    || ||  ||| | |  | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
                |   |       | | || |   || ||   ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
                |||||||| |||||| ||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
                ||||||| |||||||||   ||| || ||||||||||||||   ||||||||||||||  |
Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
                ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
                |   |||||||||||||||| | || ||||    | ||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
                |||  ||||||| |    | ||  ||| |     ||       |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745
```

```
Qy        716 TTCATTTAAGAAATTCATCTT 736
              |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 23
US-09-949-016-14565/c
; Sequence 14565, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14565
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14565

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| |||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |      ||| |   | |   | |  |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                  |   |     ||| || |  || | | |     |   |        | ||
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| |||  |   |      |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||              ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
              ||| |   ||| |||    || ||  ||| | |  | |||  | |  |  |  |    |
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419
              |    |        | | || |   || ||   ||||||| || || |||| ||||||
Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479
              |||||||| |||||| |||||||||||||||||||||||||||||||||||||| ||||
Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539
              ||||||| |||||||||  ||| || ||||||||||||||  |||||||||||||||  |

Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| |||| |||| ||||| || ||  ||||| ||||| ||||||||||||||||||
Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655
              |   |||||||||||||||| | || ||||   | ||||||||||||||| ||  |||
Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715
              |||  ||||||| |    | ||  ||| |     ||      |||  | |||   | ||
Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736
              |  |  |||  ||| | |
Db     545744 GGAAGATCTGAATCTCAGCAT 545724


RESULT 24
US-09-949-016-14566/c
; Sequence 14566, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14566
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14566

  Query Match             11.2%;  Score 223.8;  DB 3;  Length 818128;
  Best Local Similarity   61.3%;  Pred. No. 4.4e-45;
  Matches  454;  Conservative    0;  Mismatches  267;  Indels   20;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||| |||||||||||||||||||||| ||||||||| |||||||||||||| |
Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119
              | | |   ||| | || |||  ||||    |      ||| |   | |   | |  |
Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179
                  |   |     ||| || |  || | | |     |   |        | ||
Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239
              | |   |    || ||   || |||| |||  |   |      |  |  |||| ||
Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299
              ||              ||  |   | || |   | |  || |   || |   | || ||
Db     546209 TTAG------------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359
```

```
Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419

Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479

Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539

Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599

Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655

Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715

Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736

Db     545744 GGAAGATCTGAATCTCAGCAT 545724
```

RESULT 25
US-09-949-016-14567/c
; Sequence 14567, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14567
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14567

Query Match 11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4.4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60

Db     546449 TGATCCACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGAGTGAGCCACCACGCCCG 546390

Qy         61 GTCGAATGTCTTTATTATTT-GAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGA 119

Db     546389 GCCTATCATCTGTCTTCTTTACAAGAATGTAAGCTCTCCTGAGGGCAGGGGTTTTGTCTT 546330

Qy        120 CAGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTA 179

Db     546329 GTTTGCTGCTTTATTATCAATACCTAAAATGGCGCCTGGCATATAACAGACCCACAATAG 546270

Qy        180 GTCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGA 239

Db     546269 ATATTAATTGAATGACTAAGTGAATATTTAAACTCCTAGCTAAAAAGTCTTCAGGGTTTC 546210

Qy        240 ATAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTA 299

Db     546209 TTAG----------GTGTAATCTGGTTTAACACACCAAGTTGAGAATCACAGAATTA 546163

Qy        300 GTAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAA 359

Db     546162 TTAGAT--AATTAAAGCGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

Qy        360 CTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTG 419

Db     546104 AGACAGAAAAAGTGAAAGTCAAAATGGTTAAATGAGGCCAGGAGCGGTGGCTCATGCCTG 546045

Qy        420 TAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGAC 479

Db     546044 TAATCCCAGCACTTTAGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGAC 545985

Qy        480 CAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGT 539

Db     545984 CAGCCTGACCAACATGGCAAAATCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCAT 545925

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599

Db     545924 GGTGGCAGGCACTTGTAATCCCAGCTGCTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC 545865

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCT-GGCAACA 655

Db     545864 CCGTGAGGTGGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGACA 545805

Qy        656 GAGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATC 715

Db     545804 GAGTGAGACTCCGTCTCAAAAAACAAAACAAAACAAAACCAAAAAACATTTGAGACCATT 545745

Qy        716 TTCATTTAAGAAATTCATCTT 736

Db     545744 GGAAGATCTGAATCTCAGCAT 545724
```

RESULT 26
US-09-949-016-11957/c
; Sequence 11957, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11957
; LENGTH: 64813
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11957

Query Match 10.9%; Score 217.6; DB 3; Length 64813;
Best Local Similarity 59.7%; Pred. No. 6.4e-44;
Matches 426; Conservative 0; Mismatches 274; Indels 14; Gaps 3;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||  |||  ||||||||||||||||||||||||||||| ||||||||||||||
Db      57787 TGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGCCCA 57728

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  || |||  |||  ||||  || |              |   |  |   |      |
Db      57727 GCCTCATTTCTACATTTCTTGATTAGCCTTACAAATGTGCATGCCCTTTGACCCCAGAAA 57668

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | || ||  | |  ||          ||   || |          |   |  |    | |
Db      57667 TTCTACTTATAGAAATACAGGTCAGAAAAATAACCAGATGTATAAAAACATTTATGTGAC 57608

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | | |  ||  |   ||   ||     ||   | |||             ||  | |
Db      57607 TATTTACCATATTATTTAAAACAATGAAAATTGGAAATGGCACAATTGACTAACAATGGT 57548

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |      ||  |  ||    |  ||  |||||     | ||||     |||  |    |
Db      57547 TGCCTGAATAATGGGTAAATAAACTATGGTTTGTCTGTAGATAAAATACTCATCGGCCGG 57488

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||  |  ||  | ||  |  | |  ||  | ||  |  |  |  |      |
Db      57487 TAAATATCATATTATATAAGAATATCAGATAACATAAGAAAATGTTCTCGGT-------- 57436

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
                |||  |||  |    | | |  |        |||||||| |||| ||||   ||| |
Db      57435 --ATATTGTTAGGTAGAAAGAAAAATCAAGTTACGGCCAGGCACAGTGGCTCACACCTAT 57378

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||  |||  ||||||||||  |||||||||||||  |||||
Db      57377 AATCCCAGCACTTTGGGAGGCTGAGGCAGGCGGATCACTGGAGGTCAGGAGTTCGAGACC 57318

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||| ||||||||  ||||||||||  ||||||||||||||  |||||||||||||||   |
Db      57317 AGCTTGGCCAACGTGGTGAAACCCCGTCTCTACTAAAAACACAAAAATTAGCCAGGCACG 57258

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||  ||| |||||||||||||  ||||||||||  | |||||  ||||||||  |||||||||
Db      57257 GTGGTGTGCGCCTGTAGTCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCGCTTGAACC 57198

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCTGG-CAACAG 656
              | ||||||||||||||  | | | |||||   ||| ||||||||||||||||||   ||||
Db      57197 TGGGAGGTGGAGGTTTTACTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGATGACAG 57138

Qy        657 AGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTAC 710
              ||  |||||||||  || | || |||||  |   |||||  || | |     || ||
Db      57137 AGTGAGACTCCATTAAAAAAAAAAAAGTTACAAAATTATGCACAAATTATGCAC 57084
```

RESULT 27
US-09-949-016-16064/c
; Sequence 16064, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16064
; LENGTH: 70131
; TYPE: DNA
; ORGANISM: Human

US-09-949-016-16064

Query Match 10.9%; Score 217.6; DB 3; Length 70131;
Best Local Similarity 59.7%; Pred. No. 6.5e-44;
Matches 426; Conservative 0; Mismatches 274; Indels 14; Gaps 3;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              ||||||||||  |||  ||||||||||||||||||||||||||||| ||||||||||||||
Db      63441 TGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGCCCA 63382

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |  || |||  |||  ||||  || |              |   |  |   |      |
Db      63381 GCCTCATTTCTACATTTCTTGATTAGCCTTACAAATGTGCATGCCCTTTGACCCCAGAAA 63322

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | || ||  | |  ||          ||   || |          |   |  |    | |
Db      63321 TTCTACTTATAGAAATACAGGTCAGAAAAATAACCAGATGTATAAAAACATTTATGTGAC 63262

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | | |  ||  |   ||   ||     ||   | |||             ||  | |
Db      63261 TATTTACCATATTATTTAAAACAATGAAAATTGGAAATGGCACAATTGACTAACAATGGT 63202

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              |      ||  |  ||    |  ||  |||||     | ||||     |||  |    |
Db      63201 TGCCTGAATAATGGGTAAATAAACTATGGTTTGTCTGTAGATAAAATACTCATCGGCCGG 63142

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              ||  |  ||  | ||  |  | |  ||  | ||  |  |  |  |      |
Db      63141 TAAATATCATATTATATAAGAATATCAGATAACATAAGAAAATGTTCTCGGT-------- 63090

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
                |||  |||  |    | | |  |        |||||||| |||| ||||   ||| |
Db      63089 --ATATTGTTAGGTAGAAAGAAAAATCAAGTTACGGCCAGGCACAGTGGCTCACACCTAT 63032

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||  |||  ||||||||||  |||||||||||||  |||||
Db      63031 AATCCCAGCACTTTGGGAGGCTGAGGCAGGCGGATCACTGGAGGTCAGGAGTTCGAGACC 62972

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              ||| ||||||||  ||||||||||  ||||||||||||||  |||||||||||||||   |
Db      62971 AGCTTGGCCAACGTGGTGAAACCCCGTCTCTACTAAAAACACAAAAATTAGCCAGGCACG 62912

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||  ||| |||||||||||||  ||||||||||  | |||||  ||||||||  |||||||||
Db      62911 GTGGTGTGCGCCTGTAGTCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCGCTTGAACC 62852

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCTGG-CAACAG 656
              | ||||||||||||||  | | | |||||   ||| ||||||||||||||||||   ||||
Db      62851 TGGGAGGTGGAGGTTTTACTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGATGACAG 62792

Qy        657 AGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTAC 710
              ||  |||||||||  || | || |||||  |   |||||  || | |     || ||
Db      62791 AGTGAGACTCCATTAAAAAAAAAAAAGTTACAAAATTATGCACAAATTATGCAC 62738
```

RESULT 28
US-09-768-185A-1/c
; Sequence 1, Application US/09768185A
; Patent No. 6818758
; GENERAL INFORMATION:
; APPLICANT: Cassel, Michael et al
; TITLE OF INVENTION: Estrogen receptor beta variants and
; TITLE OF INVENTION: methods of detection thereof
; FILE REFERENCE: CL000280
; CURRENT APPLICATION NUMBER: US/09/768,185A
; CURRENT FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 09768185
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 325791
; TYPE: DNA

```
;   ORGANISM: HUMAN
US-09-768-185A-1

  Query Match            10.7%;  Score 214.4;  DB 3;  Length 325791;
  Best Local Similarity   73.0%;  Pred. No. 7e-43;
  Matches  305;  Conservative    0;  Mismatches  106;  Indels    7;  Gaps    2;

Qy        268 TGTTTGATATGAAATAAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATTTAAGAGTTA 327
              || |||| | || | |    | |||||||||  |        ||| || ||| ||  |
Db     257049 TGATTGAGAAGATAGATATTTATCAACAAGTCTTCAAACAAATTTAAAATTAAAATGAGG 256990

Qy        328 TTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTT 387
              |   | | |  |    ||  |||   | ||  | |   |   || |||  ||||   ||
Db     256989 ATACATTCTATGAGAGCATTTAATAGCAGAAGGTGACTTAGGAAAGTTAGAAAAGGCTTT 256930

Qy        388 TGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGT 447
                  ||||| ||||||| |||  ||| |||||||||| ||||||||||||   |||
Db     256929 ----CCTGGCCATGCGCAGTGGCTTATGCTTGTAATCCCAGCACTTTGGGAGGCTGAGGC 256874

Qy        448 GGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGT 507
              ||| ||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db     256873 GGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCTGT 256814

Qy        508 CTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTAC 567
              |||||| |||||  |||||||||||  ||| ||||||||| |||||||| ||||| ||||
Db     256813 CTCTACCAAAAATACAAAAATTAGCTGGGTATGGTGGCAGGCACCTGTAATCCCAGCTAC 256754

Qy        568 TCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGA 627
                  ||||| ||||| |||||||||||||| |||  |||||| |||||||||||||| | ||
Db     256753 CTGGGAGGCTGAGGCAGGAGAATTGCTTAAACTCAGGAGGCGGAGGTTGCAGTAAGCTGA 256694

Qy        628 GAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||   | | |||||||||||||||||| ||| |||||||||| ||    | || ||||
Db     256693 GATCACGCCCCTGCACTCCAGCCTGGCGACACAGCAAGACTCTATCTCAAAAAAAAAA 256636


RESULT 29
US-09-949-016-17578
; Sequence 17578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17578
;   LENGTH: 43726
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(43726)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17578

  Query Match            10.6%;  Score 212.2;  DB 3;  Length 43726;
  Best Local Similarity   81.0%;  Pred. No. 1.2e-42;
  Matches  260;  Conservative    0;  Mismatches   58;  Indels    3;  Gaps    1;

Qy        364 TACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAAT 423
              |||||  ||||    |  | |  | ||| |  |||| || ||||| ||||  |||| ||||
Db      23622 TACACAAAAATGTGTGGAAAAATGTGAGAGCGGCCGGGTGCAGTGGCTCACGCCTATAAT 23681

Qy        424 CCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGC 483
              || | |||||||||||||||||| ||| |||||||||||||||||||||||| || ||| |
Db      23682 CCTAGCACTTTGGGAGGCCAAGGCGGGTGGATCACCTGAGGTCAGGAGTTTGAGCCCATC 23741

Qy        484 CTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTG 543
              |||||||||||| |||||||  ||| |||||||||  ||||||||||  ||  |||||
Db      23742 CTGGCCAACATGGCGAAACCCCATCTTTACTAAAAATACAAAAATTAGCTGGGCATGGTG 23801

Qy        544 GCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAG 603
              |||||||||||||||||| |||||  |||||  | ||| |||||||||  |||||||| ||
Db      23802 GCATGCACCTGTAGTCCCAGCTACTTGGGAGGCTAAGGCAGGAGAATCACTTGAACCCAG 23861

Qy        604 GAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCA 660
              |||||||||||||||||| | |||||||    | || |||| ||||||||||||||||||||
Db      23862 GAGGTGGAGGTTGCAGTGAGCCGAGATCGCGCCATTGCATTCCAGCCTGGCAACAGAGCA 23921

Qy        661 AGACTCCATAAAGACAACAAA 681
              |||||||||     | ||  ||
Db      23922 AGACTCCATCTCAAAAAATAA 23942


RESULT 30
US-09-949-016-17482
; Sequence 17482, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17482
;   LENGTH: 76962
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(76962)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17482

  Query Match            10.6%;  Score 212.2;  DB 3;  Length 76962;
  Best Local Similarity   78.9%;  Pred. No. 1.5e-42;
  Matches  266;  Conservative    0;  Mismatches   68;  Indels    3;  Gaps    1;

Qy        378 TAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTGGG 437
              | ||||||          |||| |||||||| ||||  |||||||||||| |||||||||
Db      45917 TTAAGAGCAGACCTGGTGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 45976

Qy        438 AGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGT 497
              |||||  ||| |||||||||||||||||||| ||||||  |||||||||||||||||||| |
Db      45977 AGGCCGAGGCGGGCGGATCACCTGAGGTCAAGAGTTTCAGACCAGCCTGGCCAACATGAT 46036

Qy        498 GAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGTAG 557
              |||||||||||||||||||||    |||||||||| ||||   ||| |||| |||||||||
Db      46037 GAAACCCTGTCTCTACTAAAAGTACAAAAATTAGGCAGGCATGGCGGCAGGCACCTGTAA 46096

Qy        558 TCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTTGC 617
              |||||||||||||||||||| ||||| |||||||||||||||| |||||  |||||||||
Db      46097 TCCCAACTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGC 46156

Qy        618 AGTAACCCGAG---ATGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAAGA 674
              ||| | | |||   ||| |||||||||||||||||     |||  |||||||||     |
Db      46157 AGTGAGCTGAGATCATGCCACTGCACTCCAGCCTGGGTGACGAGAGAGACTCCATCTCAA 46216
```

```
Qy        675 CAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACC 711
              || ||||      |||       ||| |||  |  ||
Db      46217 AAAAAAAAAAAAAAAAAAAAAAAAAAAGAGCAGAGCC 46253
```

RESULT 31
US-09-949-016-12550
; Sequence 12550, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12550
; LENGTH: 45842
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12550

Query Match 10.6%; Score 211.8; DB 3; Length 45842;
Best Local Similarity 65.0%; Pred. No. 1.6e-42;
Matches 379; Conservative 0; Mismatches 192; Indels 12; Gaps 4;

```
Qy        259 TGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATT 318
              | |  | | | | |  ||  | | ||| | || ||  |   | |  | ||
Db      23739 TAAACAAAATATGTTCTAATATAAAAGTAATTAAAATATATGAATAAGATAAATTCACAA 23798

Qy        319 TAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATT--- 375
                |   || |||  ||| |  ||||||  ||    | ||  || | |||||||  |
Db      23799 ATCAAAAAATATATATCAAGTTGTTAATATGATGCCTACAAGAAACAAACTTAAAATGAT 23858

Qy        376 --ATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTT 433
                || ||| | ||   |||  |||||| |||||| ||||   | ||||||||||  |||||
Db      23859 ACATAAAAAAACTGAAAAACAGGCCAGTCGCAGTGGCTCACACTTGTAATCCCAGCACTT 23918

Qy        434 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACA 493
              ||| ||||||||||||||||||| ||||||||||||||   | ||||||||||||||
Db      23919 TGGAAGGCCAAGGTGGGCGGATCATCTGAGGTCAGGAGTTCGAAACCAGCCTGGCCAATG 23978

Qy        494 TGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCT 553
              |||||||||  ||||||||||||||||    ||||||||||||||| |||||||    || |||
Db      23979 TGGTGAAACTCTGTCTCTACTAAAAATAAAAAAATTAGCCAGGCGTGGTGGTGGGCGCCT 24038

Qy        554 GTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGG 613
                || ||||  ||||||||||||| ||||| ||||||||||||||||||| |||||  ||||
Db      24039 ATAATCCCGGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCAGAGG 24098

Qy        614 TTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCAT 669
              |||||||  | |||||    ||| |||||||||||||| ||| ||||||||||  || |
Db      24099 TTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGGCGACAGAGCAAGATGCCGT 24158

Qy        670 ---AAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGA 726
                 ||| |||  || | |    |||   || ||| | || |  |  |  |    |
Db      24159 CTCAAAAACAGGAACAACAACAAAAACCTGAAAAGCAATTCTGTTTTTTATTTCTGGGAA 24218

Qy        727 AATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTTAATTAA 786
              |  |  |   | ||      |   | ||  ||   || | ||     | || |  ||| |||
Db      24219 AGCTGCTGGCTCTTTTCCCAAGATCACAAGAGTAGACAACTATAGGAAAATAATAAATAA 24278

Qy        787 GCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATT 829
              |||| |  ||||    ||      |  || |  | |||   ||
Db      24279 ACCCTTATGGATTCGTGGTGGGGAGGGATATTTTGAACTTGTT 24321
```

RESULT 32
US-09-949-016-17327
; Sequence 17327, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17327
; LENGTH: 45842
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17327

Query Match 10.6%; Score 211.8; DB 3; Length 45842;
Best Local Similarity 65.0%; Pred. No. 1.6e-42;
Matches 379; Conservative 0; Mismatches 192; Indels 12; Gaps 4;

```
Qy        259 TGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAGTAGCTGCTATTGAAGATT 318
              | |  | | | | |  ||  | | ||| | || ||  |   | |  | ||
Db      23739 TAAACAAAATATGTTCTAATATAAAAGTAATTAAAATATATGAATAAGATAAATTCACAA 23798

Qy        319 TAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAACTAATACACTTAAATT--- 375
                |   || |||  ||| |  ||||||  ||    | ||  || | |||||||  |
Db      23799 ATCAAAAAATATATATCAAGTTGTTAATATGATGCCTACAAGAAACAAACTTAAAATGAT 23858

Qy        376 --ATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTT 433
                || ||| | ||   |||  |||||| |||||| ||||   | ||||||||||  |||||
Db      23859 ACATAAAAAAACTGAAAAACAGGCCAGTCGCAGTGGCTCACACTTGTAATCCCAGCACTT 23918

Qy        434 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACA 493
              ||| ||||||||||||||||||| ||||||||||||||   | ||||||||||||||
Db      23919 TGGAAGGCCAAGGTGGGCGGATCATCTGAGGTCAGGAGTTCGAAACCAGCCTGGCCAATG 23978

Qy        494 TGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCT 553
              |||||||||  ||||||||||||||||    ||||||||||||||| |||||||    || |||
Db      23979 TGGTGAAACTCTGTCTCTACTAAAAATAAAAAAATTAGCCAGGCGTGGTGGTGGGCGCCT 24038

Qy        554 GTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGG 613
                || ||||  ||||||||||||| ||||| ||||||||||||||||||| |||||  ||||
Db      24039 ATAATCCCGGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCAGAGG 24098

Qy        614 TTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCCAT 669
              |||||||  | |||||    ||| |||||||||||||| ||| ||||||||||  || |
Db      24099 TTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGGCGACAGAGCAAGATGCCGT 24158

Qy        670 ---AAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCTATCTTCATTTAAGA 726
                 ||| |||  || | |    |||   || ||| | || |  |  |  |    |
Db      24159 CTCAAAAACAGGAACAACAACAAAAACCTGAAAAGCAATTCTGTTTTTTATTTCTGGGAA 24218

Qy        727 AATTCATCTTTGTTCATTTATTTTTACTTGACATGAGAGCTTCCAGCAATTTTTAATTAA 786
              |  |  |   | ||      |   | ||  ||   || | ||     | || |  ||| |||
Db      24219 AGCTGCTGGCTCTTTTCCCAAGATCACAAGAGTAGACAACTATAGGAAAATAATAAATAA 24278

Qy        787 GCCCTCACAGATTTTATGTCACTGGCTATGTGATAAACAAATT 829
              |||| |  ||||    ||      |  || |  | |||   ||
Db      24279 ACCCTTATGGATTCGTGGTGGGGAGGGATATTTTGAACTTGTT 24321
```

```
RESULT 33
US-09-949-016-12731/c
; Sequence 12731, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12731
;   LENGTH: 264206
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12731

  Query Match             10.6%;  Score 211.6;  DB 3;  Length 264206;
  Best Local Similarity   80.1%;  Pred. No. 3.2e-42;
  Matches  274;  Conservative    0;  Mismatches   64;  Indels    4;  Gaps    2;

Qy        372 AATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACAC 431
              || | ||||||  | |   ||| |||| ||  |||| ||||   |||||||||||| |||
Db      60222 AAATCTTAAAGTCCATCTTAATAGGCCGGGGACAGTGGCTCACACCTGTAATCCCAGCAC 60163

Qy        432 TTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAA 491
              |||||||||||||||| |||||||||||||||| |||||||||  ||||||||||||||||
Db      60162 TTTGGGAGGCCAAGGCGGGCGGATCACCTGAGATCAGGAGTTCGAGACCAGCCTGGCCAA 60103

Qy        492 CATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCAC 551
              |||||||||||||||| |||| ||||||  |||||||||||||| |||||||||| | |
Db      60102 CATGGTGAAACCCTGTGTCTAATAAAAATACAAAAATTAGCCAGATGTGGTGGCAGGTGC 60043

Qy        552 CTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGA 611
              ||||| ||||| |||||| ||||| |||||  ||||||| |||||| ||| || || |||
Db      60042 CTGTAATCCCAGCTACTCGGGAGGCTGAGGTGGGAGAATCGCTTGAGCCTGGGGGGCGGA 59983

Qy        612 GGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCC 667
              ||||||||| | |||||    ||| |||||||||||||| |||||||||||||||||||
Db      59982 GGTTGCAGTGAGCCGAGACCATGCCACTGCACTCCAGCCTGGGCAACAGAGCAAGACTCC 59923

Qy        668 ATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTA 709
              |||   | || ||||      |     | | ||| | ||| |
Db      59922 ATATCAAAAAAAAAAAAAAAAAGTCCATCTTAATAAATTGGA 59881


RESULT 34
US-09-949-016-13249/c
; Sequence 13249, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13249
;   LENGTH: 264304
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13249

  Query Match             10.6%;  Score 211.6;  DB 3;  Length 264304;
  Best Local Similarity   80.1%;  Pred. No. 3.2e-42;
  Matches  274;  Conservative    0;  Mismatches   64;  Indels    4;  Gaps    2;

Qy        372 AATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACAC 431
              || | ||||||  | |   ||| |||| ||  |||| ||||   |||||||||||| |||
Db      60222 AAATCTTAAAGTCCATCTTAATAGGCCGGGGACAGTGGCTCACACCTGTAATCCCAGCAC 60163

Qy        432 TTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAA 491
              |||||||||||||||| |||||||||||||||| |||||||||  ||||||||||||||||
Db      60162 TTTGGGAGGCCAAGGCGGGCGGATCACCTGAGATCAGGAGTTCGAGACCAGCCTGGCCAA 60103

Qy        492 CATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCAC 551
              |||||||||||||||| |||| ||||||  |||||||||||||| |||||||||| | |
Db      60102 CATGGTGAAACCCTGTGTCTAATAAAAATACAAAAATTAGCCAGATGTGGTGGCAGGTGC 60043

Qy        552 CTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGA 611
              ||||| ||||| |||||| ||||| |||||  ||||||| |||||| ||| || || |||
Db      60042 CTGTAATCCCAGCTACTCGGGAGGCTGAGGTGGGAGAATCGCTTGAGCCTGGGGGGCGGA 59983

Qy        612 GGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACAGAGCAAGACTCC 667
              ||||||||| | |||||    ||| |||||||||||||| |||||||||||||||||||
Db      59982 GGTTGCAGTGAGCCGAGACCATGCCACTGCACTCCAGCCTGGGCAACAGAGCAAGACTCC 59923

Qy        668 ATAAAGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTA 709
              |||   | || ||||      |     | | ||| | ||| |
Db      59922 ATATCAAAAAAAAAAAAAAAAAGTCCATCTTAATAAATTGGA 59881


RESULT 35
US-09-949-002-837
; Sequence 837, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 837
;   LENGTH: 54286
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-002-837

  Query Match             10.6%;  Score 211.4;  DB 3;  Length 54286;
  Best Local Similarity   78.3%;  Pred. No. 2.1e-42;
  Matches  267;  Conservative    0;  Mismatches   71;  Indels    3;  Gaps    1;

Qy        375 TATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTT 434
              | || || | |  |   ||||||  |||||||| |||| ||||||||||||| ||||||
Db      32756 TCTTCAAAAACACTTGTATGGGCTGGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTT 32815

Qy        435 GGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACAT 494
              |||||||  ||| |||||||||||| |||||||||||||| |||||  ||| ||||||||||
Db      32816 GGGAGGCTGAGGCGGGCGGATCACTTGAGGTCAGGAGTTCAAGACTAGCTTGGCCAACAT 32875

Qy        495 GGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTG 554
              ||| |||||||||| ||||||||||  ||||||||||| |    ||||||  |  ||||
Db      32876 GGTAAAACCCTGTCCCTACTAAAAATACAAAAATTAGCCGGACAGGGTGGCGGGTGCCTG 32935
```

```
Qy        555 TAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGT 614
              |||||||| |||||||||||| ||||| |||||||| | ||||||| ||||||  |||||
Db      32936 TAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGTTTGAACCCAGGAGGCAGAGGT 32995

Qy        615 TGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAA 671
              |||||| | ||| |   ||  ||||||||||||||||||||||||||| |||| ||||
Db      32996 TGCAGTGAGCCGCGATCATCCCACTGCACTCCAGCCTGGCAACAGAGCGAGACACCATCT 33055

Qy        672 AGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCT 712
              | | || ||||       |    | |  ||| ||| | |||
Db      33056 AAAAAAAAAAAAAACAACACACATTTGTATGGGTTTTCCCT 33096


RESULT 36
US-09-949-002-642
; Sequence 642, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 642
;   LENGTH: 66955
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-002-642

  Query Match             10.6%;  Score 211.4;  DB 3;  Length 66955;
  Best Local Similarity   78.3%;  Pred. No. 2.3e-42;
  Matches  267;  Conservative    0;  Mismatches   71;  Indels    3;  Gaps    1;

Qy        375 TATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTT 434
              | || || | |  |    |||||| |||||||| |||| ||||||||||||||| ||||||
Db      45297 TCTTCAAAAACACTTGTATGGGCTGGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTT 45356

Qy        435 GGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACAT 494
              |||||||  ||| ||||||||||| |||||||||||||| ||||| ||| ||||||||||
Db      45357 GGGAGGCTGAGGCGGGCGGATCACTTGAGGTCAGGAGTTCAAGACTAGCTTGGCCAACAT 45416

Qy        495 GGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTG 554
              ||| |||||||||| |||||||||  |||||||||||  |    ||||||  |  ||||
Db      45417 GGTAAAACCCTGTCCCTACTAAAAATACAAAAATTAGCCGGACAGGGTGGCGGGTGCCTG 45476

Qy        555 TAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGT 614
              |||||||| |||||||||||| ||||| |||||||| | ||||||| ||||||  |||||
Db      45477 TAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGTTTGAACCCAGGAGGCAGAGGT 45536

Qy        615 TGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAA 671
              |||||| | ||| |   ||  ||||||||||||||||||||||||||| |||| ||||
Db      45537 TGCAGTGAGCCGCGATCATCCCACTGCACTCCAGCCTGGCAACAGAGCGAGACACCATCT 45596

Qy        672 AGACAACAAAAGCTTTGAAATTGTGTAAATGAGTTGTACCT 712
              | | || ||||       |    | |  ||| ||| | |||
Db      45597 AAAAAAAAAAAAAACAACACACATTTGTATGGGTTTTCCCT 45637


RESULT 37
US-09-949-016-17270/c
; Sequence 17270, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17270
;   LENGTH: 12579
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(12579)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17270

  Query Match             10.5%;  Score 210.8;  DB 3;  Length 12579;
  Best Local Similarity   58.1%;  Pred. No. 1.8e-42;
  Matches  410;  Conservative    0;  Mismatches  292;  Indels    4;  Gaps    2;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              | |||| ||  |||||||||||||||||||||||||||||||||||||||||  | ||||
Db       7608 TCATCCTCCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACGGCTCCTG 7549

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              |||  |  |   | ||| ||| | |          |   ||  |  |||  |  || ||
Db       7548 GTCTGAATTAAATTTTACTTGGAAAATTTCTTGAGAATTTTCACATGTAATACATATGCA 7489

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
              | || |    | ||  | |  | ||          | |   || |         |||   ||
Db       7488 AAACATCCCAGAAGGGACAGACAAAATTCCTGCCATGAAGACCTTGTGGTGATGTTGAAG 7429

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
                          |     |||  || ||  | |         ||| |
Db       7428 AGAAAGGACATGTGTATATGTATGAAATAATTTGAAAACAAGTCCAAGACAAAGTAAACG 7369

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
                  | |||   ||       | | | ||| |  |  |    || |   |    |
Db       7368 AGGGGAAGAGGAAACGCTTTGATTGAATATATGAGACCAGCTCTTCATATACCCTCACAC 7309

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              | |  ||       |       ||| |||       |  | |  ||   | | |
Db       7308 TCACATATAGGTGGACTCCTGTGTTTACTTACCGGTTCCTTAAACATAATAAGTTAGATT 7249

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
                |    |||| |   || |       |  |||||||||||||| |||| ||||| |
Db       7248 AGCAAGCTCTAAAGTCAGAATGTTTGAAAAGCAGGGCCAGGCGCAGTGGCTCATGCCTAT 7189

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||| ||| |||| |||||| ||||| |||||||||| |||||
Db       7188 AATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCACTTGAGGCCAGGAGTTTGAGACC 7129

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||||||||||||| ||||||  ||||  ||| |||| ||||||||||  ||  ||
Db       7128 AGCCTGGCCAACATGGCGAAACCACGTCTTCACTGAAAATACAAAAATTAGCTGGGCATG 7069

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              |||||  |||||||| ||||| |||   |||||| ||||| |||||||||||||| ||||
Db       7068 GTGGCCGGCACCTGTAATCCCAGCTATATAGGAGGCTGAGGCAGGAGAATTGCTTAAACC 7009

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCC-TGGCAACAG 656
              |||||  || |||||||| | || |||   || ||||||||||||||  |||||||
Db       7008 CGGGAGGCAGAAGTTGCAGTGAGCCAAGATTGTGCCACTGCACTCCAGCCAGGGCAACAG 6949

Qy        657 AGCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAATG 702
              ||   |||||  |    | || ||||   |   |||  |  ||| |
Db       6948 AGGCCGACTCTGTCCCAAAAAAAAAAAAGTAAGAATGTTCGAAAAG 6903
```

```
RESULT 38
US-09-949-016-12682/c
; Sequence 12682, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12682
;   LENGTH: 63187
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12682

  Query Match             10.5%;  Score 210.6;  DB 3;  Length 63187;
  Best Local Similarity   78.6%;  Pred. No. 3.5e-42;
  Matches  265;  Conservative    0;  Mismatches   69;  Indels    3;  Gaps    1;

Qy        349 AATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTA 408
              || || |  |  ||| ||  | ||| ||   || |      |  | ||||||||  ||
Db      35113 AACTTAATTAGATAACACCATGAAAATAAATAAAAATGCACACTTTGGCCAGGCATGGTG 35054

Qy        409 GCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAG 468
              |||| || || |||||||| |||||||||||||||||||||||| ||||||||||||||||
Db      35053 GCTCATGTCTATAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAGATCACCTGAGGTCAG 34994

Qy        469 GAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAAT 528
              ||||| ||||||| |||||||||||||| ||| ||||||||||||| |||  |||||||
Db      34993 GAGTTCAAGACCATCCTGGCCAACATGGTAAAATCCTGTCTCTACTACAAATACAAAAAT 34934

Qy        529 TAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGA 588
              ||| |||||||| |||||  |||||||| | | | |||||  ||||| ||||| | ||||
Db      34933 TAGGCAGGTGTGATGGCACACACCTGTAATTCTAGCTACTGGGGAGGCTGAGGCAAGAGA 34874

Qy        589 ATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAG 645
              |||||||||||| ||||| |||||||||||  | || ||||   | ||||||||||||
Db      34873 ATTGCTTGAACCTGGGAGGGGGAGGTTGCAGGGAGCCAAGATCACGCCACTGCACTCCAG 34814

Qy        646 CCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||| ||||||| |||||||||    | || || |
Db      34813 CCTGGCCACAGAGCGAGACTCCATCTCAAAAAAAATA 34777


RESULT 39
US-09-949-016-16288/c
; Sequence 16288, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16288
;   LENGTH: 63187
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16288

  Query Match             10.5%;  Score 210.6;  DB 3;  Length 63187;
  Best Local Similarity   78.6%;  Pred. No. 3.5e-42;
  Matches  265;  Conservative    0;  Mismatches   69;  Indels    3;  Gaps    1;

Qy        349 AATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTA 408
              || || |  |  ||| ||  | ||| ||   || |      |  | ||||||||  ||
Db      35113 AACTTAATTAGATAACACCATGAAAATAAATAAAAATGCACACTTTGGCCAGGCATGGTG 35054

Qy        409 GCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAG 468
              |||| || || |||||||| |||||||||||||||||||||||| ||||||||||||||||
Db      35053 GCTCATGTCTATAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAGATCACCTGAGGTCAG 34994

Qy        469 GAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAAT 528
              ||||| ||||||| |||||||||||||| ||| ||||||||||||| |||  |||||||
Db      34993 GAGTTCAAGACCATCCTGGCCAACATGGTAAAATCCTGTCTCTACTACAAATACAAAAAT 34934

Qy        529 TAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGA 588
              ||| |||||||| |||||  |||||||| | | | |||||  ||||| ||||| | ||||
Db      34933 TAGGCAGGTGTGATGGCACACACCTGTAATTCTAGCTACTGGGGAGGCTGAGGCAAGAGA 34874

Qy        589 ATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTCCAG 645
              |||||||||||| ||||| |||||||||||  | || ||||   | ||||||||||||
Db      34873 ATTGCTTGAACCTGGGAGGGGGAGGTTGCAGGGAGCCAAGATCACGCCACTGCACTCCAG 34814

Qy        646 CCTGGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||| ||||||| |||||||||    | || || |
Db      34813 CCTGGCCACAGAGCGAGACTCCATCTCAAAAAAAATA 34777


RESULT 40
US-09-949-016-13713/c
; Sequence 13713, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13713
;   LENGTH: 177669
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(177669)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13713

  Query Match             10.5%;  Score 210.6;  DB 3;  Length 177669;
  Best Local Similarity   80.1%;  Pred. No. 5e-42;
  Matches  273;  Conservative    0;  Mismatches   64;  Indels    4;  Gaps    2;

Qy        346 TAAAATTTTAAAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCA 405
              | | || ||   | |  |||||||||||| | |  | ||| ||| |    ||||||||
Db      90517 TCACATCTTGGTATCCCATACACTTAAAGTGTAGATGAGATTTAATGGAGGCCAGGCATG 90458

Qy        406 GTAGCTCCTGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGT 465
```

```
              || |||| |||||||||||||| |||||||||||||||||| ||| || |||| ||||||
Db      90457 GTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGGTCACTTGAGGT 90398

Qy        466 CAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAA 525
              ||||||||| |||||||||||||||||||||||||||||||| |||||||||||||  ||||
Db      90397 CAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 90338

Qy        526 AATTAGCCAGGTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGG 585
              |||||||| |  |||||||||  |||||||| ||||| |||||  ||||| ||||| | |
Db      90337 AATTAGCCTGACGTGGTGGCACACACCTGTAATCCCAGCTACTAGGGAGGCTGAGGCAAG 90278

Qy        586 AGAATTGCTTGAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGAT---GTCACTGCACTC 642
               |||||||||||||| ||| || |||||||||||| | || ||||   | | ||||||||
Db      90277 GGAATTGCTTGAACCAAGGGGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCGCTGCACTC 90218

Qy        643 CAGCCT-GGCAACAGAGCAAGACTCCATAAAGACAACAAAA 682
              |||||| ||||||||||  ||||||  |   || || ||||
Db      90217 CAGCCTGGGCAACAGAGTGAGACTCTGTCTCGAAAAAAAAA 90177
```

RESULT 41
US-09-949-016-15924
; Sequence 15924, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15924
; LENGTH: 42053
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(42053)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15924

Query Match 10.5%; Score 210.4; DB 3; Length 42053;
Best Local Similarity 80.2%; Pred. No. 3.4e-42;
Matches 260; Conservative 0; Mismatches 61; Indels 3; Gaps 1;

```
Qy        376 ATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGTAATCCCAACACTTTG 435
              | | |||||    | |  ||||| |||| || |||| |||||| |||||||||||| ||
Db      12080 AATGAAGAGTCAGGGAGAGGGCCGGGCGTGGTGGCTCATGCCTGAAATCCCAACACTGTG 12139

Qy        436 GGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATG 495
              ||||||||||  || | |||||||||||||||||| |||| ||||||||||||||||||
Db      12140 GGAGGCCAAGGCAGGTGAATCACCTGAGGTCAGGAGTTCAAGATCAGCCTGGCCAACATG 12199

Qy        496 GTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTGGTGGCATGCACCTGT 555
              ||||||||| ||||| ||||||||  |||||||||||  || |||||||| ||||||||
Db      12200 GTGAAACCCGGTCTCCACTAAAAATACAAAAATTAGCTGGGCGTGGTGGCAGGCACCTGT 12259

Qy        556 AGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACCTAGGAGGTGGAGGTT 615
              ||||| | || ||||||||| ||||| |||||||||  |||||||| ||||||  ||||||
Db      12260 AGTCCAAGCTGCTCAGGAGGGTGAGGCAGGAGAATCACTTGAACCCAGGAGGCAGAGGTT 12319

Qy        616 GCAGTAACCCGAGAT---GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCATAAA 672
              ||||| | | |||||    |  ||||||||||||||||||||||||||||| ||||||| |
Db      12320 GCAGTGAGCTGAGATTGCGCTACTGCACTCCAGCCTGGCAACAGAGCGAGACTCCGTCTT 12379

Qy        673 GACAACAAAAGCTTTGAAATTGTG 696
              | || ||||  |  |  |  |||
Db      12380 AAAAAAAAAAAATCAGGCAGAGTG 12403
```

RESULT 42
US-09-949-016-13506/c
; Sequence 13506, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13506
; LENGTH: 283538
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(283538)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13506

Query Match 10.5%; Score 210.2; DB 3; Length 283538;
Best Local Similarity 77.8%; Pred. No. 7.4e-42;
Matches 267; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

```
Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              | |  | |  | ||| | ||| |       ||  |||||||||| || ||||||    ||||
Db     137661 TCAGGCCCACATATTTTAAAATAAAGCAAAATGGGGCCAGGCACAATAGCTCACAACTGT 137602

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||| ||| ||||||| |||||||||| ||||||||||  |||||
Db     137601 AATCCCAGCACTTTGGGAAGCCGAGGTGGGTGGATCACCTGTGGTCAGGAGTTCGAGACC 137542

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAGGTGTG 540
              |||||| |||||||||||||||||||||||||||||||  ||||||||||||||||||
Db     137541 AGCCTGACCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGTGTG 137482

Qy        541 GTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAACC 600
              ||||||    | | || ||||| |||||| ||||| ||||| |||||||| |||||||||
Db     137481 GTGGCACATGCATATAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 137422

Qy        601 TAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGCAACAGA 657
              ||||||||||||||||||| | | ||||   || ||||||||||||||||||||||||
Db     137421 CTGGAGGTGGAGGTTGCAGTGAGCTGAGACTGTGCCACTGCACTCCAGCCTGGCAACAGA 137362

Qy        658 GCAAGACTCCATAAAGACAACAAAAGCTTTGAAATTGTGTAAA 700
              || ||||||| |    | |   |    |  |||| |  || |
Db     137361 GCGAGACTCCGTCTCAATCAATCATCAATCAAAATGGGATAGA 137319
```

RESULT 43
US-09-949-016-16790/c
; Sequence 16790, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016

```
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16790
;   LENGTH: 78630
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16790

  Query Match            10.5%;  Score 210;  DB 3;  Length 78630;
  Best Local Similarity   83.0%;  Pred. No. 5.3e-42;
  Matches  264;  Conservative    0;  Mismatches   50;  Indels    4;  Gaps    2;

Qy        356 AAAACTAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTG 415
              ||  | ||   | ||||   | ||||      |    ||||| || || || ||||  |
Db      60827 AACTCAAAGGAAATTAAGAAACTAAATAATGAGGCTCAGGGCCGGGTGCGGTGGCTCACG 60768

Qy        416 CCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTA 475
              ||| |||||||| |||||||||||||| | ||||||||||||||||||| ||||||| |
Db      60767 CCTATAATCCCAGCACTTTGGGAGGCCGAAGTGGGCGGATCACCTGAGGTAAGGAGTTCA 60708

Qy        476 AGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGCAAAAATTAGCCAG 535
              ||| |||||||| |||||||||||||||||||||||||||||||  ||||||||||||||
Db      60707 AGAACAGCCTGTCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAG 60648

Qy        536 GTGTGGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTT 595
              |||||||||||| || |||||||||||| |||||||||||| ||||| ||||||||||||
Db      60647 GTGTGGTGGCACGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTT 60588

Qy        596 GAACCTAGGAGGTGGAGGTTGCAGTAACCCGAGA---TGTCACTGCACTCCAGCCTGGCA 652
              |||||| |||||||||||| |||||| | || |||   || ||||||||||||||||| |
Db      60587 GAACCTGGGAGGTGGAGGCTGCAGTGAGCCAAGATCGTGCCACTGCACTCCAGCCTGGGA 60528

Qy        653 -ACAGAGCAAGACTCCAT 669
               |||||||| ||||||||
Db      60527 GACAGAGCACGACTCCAT 60510


RESULT 44
US-09-949-016-12396/c
; Sequence 12396, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USE$ THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12396
;   LENGTH: 78846
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12396

  Query Match            10.5%;  Score 209.4;  DB 3;  Length 78846;
  Best Local Similarity   61.4%;  Pred. No. 7.5e-42;
  Matches  422;  Conservative    0;  Mismatches  256;  Indels    9;  Gaps    5;

Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||| || |||||||||||||||| |||||||||||||||||||||||||||| ||||
Db      37780 TGATCCGCCTGCCTTGGCCTCCCAAAGAGCTGGGATTACAGGCGTGAGCCACCACACCTG 37721

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |     |  |   |  ||   |  | | | |    |||| |        | |  |
Db      37720 GCCTCCACTAATGTCTTATTACTGCCAGCAGAAGATTTTTAATCCATGGCATTTTCAGGT 37661

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
                || ||    |    |||| |||      |           || ||     |
Db      37660 TTCAATTAATATTTTATTGCACCAAAGCTATTTACAAAACGCACAAGTTCAAAAGTCAAA 37601

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
                | ||    ||         ||   | | |||   ||| || |  ||  ||||   |
Db      37600 AAGTACCCAATAATCAGGTAATCTACAAATGCTGGG--AATCATTGTCTTGCAGGAATCA 37543

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              | |  |||||  ||   |      |   || || |  |  |  ||| | ||  |   ||
Db      37542 TGATGCTATGGGAATCATGTTTAAATATGATTCAGTTACACAAAGCAGTTACAAGGGTTA 37483

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |   |       |  | || | |    |      ||  | |   |  || ||  | |   |
Db      37482 TCTTTAGCCAGGCTGCTTGAGGGCACAAGCCACACTGCCAGCAAGAAATATGCTCAGGCC 37423

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
              |  ||      | || | ||          ||| | ||| |||  || |||| |||||||
Db      37422 CTACCCAAGAGCAGTAGTTAAAGAAGCACAAAGGAGCCGGGCGTGGTGGCTCATGCCTGT 37363

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||| |||||||||||||  |||||||||| || ||||
Db      37362 AATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCA--TGAGGTCAGGAGATTGAGACT 37305

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGC-AAAAATTAGCCAGGTGT 539
              | |||||| |||| |||||||||| |||||||||||||  | ||||||||||| |||||
Db      37304 ATCCTGGCTAACAAGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGTGT 37245

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| || |||||||||||||| |||||  ||||| ||||| |||||||| || |||||
Db      37244 GGTGGCAGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATAGCATGAAC 37185

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACA 655
              |  |||||  ||| |||||||| | |||||   ||| |||||||||||||||| |||||||
Db      37184 CCGGGAGGCAGAGCTTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGGCAACA 37125

Qy        656 GAGCAAGACTCCATAAAGACAACAAAA 682
              |||| ||||||| |       || ||||
Db      37124 GAGCGAGACTCCGTCTCAGAAAAAAAA 37098


RESULT 45
US-09-949-016-12791/c
; Sequence 12791, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12791
;   LENGTH: 78846
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12791
```

Query Match 10.5%; Score 209.4; DB 3; Length 78846;
Best Local Similarity 61.4%; Pred. No. 7.5e-42;
Matches 422; Conservative 0; Mismatches 256; Indels 9; Gaps 5;

```
Qy          1 TGATCCACCAGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGCCTG 60
              |||||| || |||||||||||||||||| |||||||||||||||||||||||||| ||||
Db      37780 TGATCCGCCTGCCTTGGCCTCCCAAAGAGCTGGGATTACAGGCGTGAGCCACCACACCTG 37721

Qy         61 GTCGAATGTCTTTATTATTTGAAGAGACAACATGGGCCTTAAATCTGTCTTCTATTTGAC 120
              | |     |  |   |  ||   |  |  | | |    ||||  |       | |  |
Db      37720 GCCTCCACTAATGTCTTATTACTGCCAGCAGAAGATTTTTAATCCATGGCATTTTCAGGT 37661

Qy        121 AGACTTTGATGGAGTCAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTAG 180
                || ||    |      |||| |||       |           || ||      |
Db      37660 TTCAATTAATATTTTATTGCACCAAAGCTATTTACAAAACGCACAAGTTCAAAAGTCAAA 37601

Qy        181 TCTCTCTCAGCTGTCTTTCTGCATATGTAAGGTGGAATAATGATGGCTTTCAAGGAGGAA 240
              | ||    ||        ||   | | |||   ||| || |  ||  ||||   |
Db      37600 AAGTACCCAATAATCAGGTAATCTACAAATGCTGGG--AATCATTGTCTTGCAGGAATCA 37543

Qy        241 TAAACCTATGAAAAGTGTTGAGGATAGTGTTTGATATGAAATAAGGATTTCAACAAGTAG 300
              | |  |||||  ||   |      |   || || |  |  |  ||| | ||  |   ||
Db      37542 TGATGCTATGGGAATCATGTTTAAATATGATTCAGTTACACAAAGCAGTTACAAGGGTTA 37483

Qy        301 TAGCTGCTATTGAAGATTTAAGAGTTATTTATTACAACTATTTAATAAAATTTTAAAAAC 360
              |  |       |  | || | |      |       ||  | |   |  || ||  | |   |
Db      37482 TCTTTAGCCAGGCTGCTTGAGGGCACAAGCCACACTGCCAGCAAGAAATATGCTCAGGCC 37423

Qy        361 TAATACACTTAAATTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCCTGCCTGT 420
                |  ||      | || | ||        ||| | ||| ||||  || |||| |||||||
Db      37422 CTACCCAAGAGCAGTAGTTAAAGAAGCACAAAGGAGCCGGGCGTGGTGGCTCATGCCTGT 37363

Qy        421 AATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTAAGACC 480
              ||||||| |||||||||||||| |||||||||||||||  |||||||||||| || ||||
Db      37362 AATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCA--TGAGGTCAGGAGATTGAGACT 37305

Qy        481 AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAACGC-AAAAATTAGCCAGGTGT 539
              | |||||| |||| |||||||||| |||||||||||||  | ||||||||||| |||||
Db      37304 ATCCTGGCTAACAAGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGTGT 37245

Qy        540 GGTGGCATGCACCTGTAGTCCCAACTACTCAGGAGGTTGAGGGAGGAGAATTGCTTGAAC 599
              ||||||| || |||||||||||| |||||  |||||  ||||| |||||||| || |||||
Db      37244 GGTGGCAGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATAGCATGAAC 37185

Qy        600 CTAGGAGGTGGAGGTTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCAACA 655
              |  |||||  ||| |||||||  | |||||    ||| ||||||||||||||| |||||||
Db      37184 CCGGGAGGCAGAGCTTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGGCAACA 37125

Qy        656 GAGCAAGACTCCATAAAGACAACAAAA 682
              |||| ||||||| |     || ||||
Db      37124 GAGCGAGACTCCGTCTCAGAAAAAAAA 37098
```

Search completed: January 22, 2006, 21:30:11
Job time : 280.224 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:10:03 ; Search time 5419.54 Seconds
(without alignments)
17274.699 Million cell updates/sec

Title: US-09-728-552A-3_COPY_59000_61000
Perfect score: 2001
Sequence: 1 aaaggcttctgaaatcagac.....cttgaagggttagagggact 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	592.4	29.6	717	10	AG115127	AG115127 Pan trogl
	2	504.6	25.2	855	3	BM041081	BM041081 603614811
c	3	476.8	23.8	836	9	BZ608243	BZ608243 WHACT79TR
	4	458	22.9	655	10	AG045844	AG045844 Pan trogl
	5	456.6	22.8	672	7	CR745511	CR745511 CR745511
c	6	450.8	22.5	860	9	AQ745404	AQ745404 HS_2276_A
	7	445.8	22.3	680	10	AG169322	AG169322 Pan trogl
c	8	440.8	22.0	649	5	BX507549	BX507549 DKFZp686I
	9	429.8	21.5	621	9	B72370	B72370 RPCI11-8J21
	10	424.2	21.2	940	5	BQ677568	BQ677568 AGENCOURT
c	11	421.6	21.1	3739	4	HSM805177	AL833881 Homo sapi
	12	415.6	20.8	577	7	CR537266	CR537266 DKFZp459N
	13	411	20.5	746	6	CF597035	CF597035 AGENCOURT
c	14	408.8	20.4	630	9	AQ390354	AQ390354 CITBI-E1-
c	15	401.6	20.1	899	5	BQ926428	BQ926428 AGENCOURT
	16	397.2	19.9	900	5	BQ678014	BQ678014 AGENCOURT
	17	392.8	19.6	816	5	BU602065	BU602065 AGENCOURT
c	18	390.6	19.5	635	6	CD685118	CD685118 EST1638 h
c	19	389.4	19.5	617	1	AL704833	AL704833 DKFZp686B
	20	387.8	19.4	741	10	AG030015	AG030015 Pan trogl
	21	386	19.3	650	10	AG085080	AG085080 Pan trogl
c	22	376.4	18.8	846	6	CD171662	CD171662 AGENCOURT
	23	370	18.5	929	6	CD242806	CD242806 AGENCOURT
c	24	369.6	18.5	821	9	BZ600660	BZ600660 WHABB74TF
	25	369.4	18.5	870	3	BM041043	BM041043 603614780
	26	369.2	18.5	937	5	BQ680536	BQ680536 AGENCOURT
	27	367	18.3	889	5	BU168299	BU168299 AGENCOURT
c	28	365	18.2	555	9	AQ356187	AQ356187 CITBI-E1-
	29	365	18.2	789	6	CF785943	CF785943 AGENCOURT
	30	363	18.1	14771	9	AQ839854	AQ839854 260L13-C5
c	31	362.2	18.1	764	9	AQ742559	AQ742559 HS_5384_B
	32	357.2	17.9	732	6	CB984929	CB984929 AGENCOURT
	33	355.4	17.8	675	10	AG132879	AG132879 Pan trogl
c	34	354.8	17.7	636	1	AL603245	AL603245 DKFZp686N
c	35	347.2	17.4	845	9	AQ747612	AQ747612 HS_5533_A
	36	346.4	17.3	874	5	BU189030	BU189030 AGENCOURT
c	37	345.8	17.3	607	9	AQ587990	AQ587990 CITBI-E1-
c	38	341.4	17.1	680	9	AQ239957	AQ239957 CIT-HSP-2
c	39	340.6	17.0	566	9	B66172	B66172 CIT-HSP-202
c	40	340.4	17.0	497	9	AQ481794	AQ481794 RPCI-11-2
	41	338.4	16.9	666	2	BE393102	BE393102 601309151
c	42	338.2	16.9	424	1	AI276815	AI276815 qk61d07.x
c	43	337.4	16.9	681	10	AG113408	AG113408 Pan trogl
	44	333.8	16.7	600	9	AQ083508	AQ083508 RPCI11-52
c	45	333.2	16.7	556	9	AQ507503	AQ507503 RPCI-11-2
c	46	326.6	16.3	652	3	BM978961	BM978961 UI-CF-DU1
	47	325.6	16.3	711	9	AQ057655	AQ057655 CIT-HSP-2
c	48	325	16.2	649	9	BZ606199	BZ606199 WHACE54TF
c	49	323.6	16.2	658	5	BX487157	BX487157 DKFZp686I
	50	322.6	16.1	476	2	BE395424	BE395424 601309938
c	51	321.8	16.1	583	10	AG019606	AG019606 Homo sapi
	52	317.4	15.9	944	2	BG334280	BG334280 602461784
c	53	316.4	15.8	1914	4	AF289565	AF289565 Homo sapi
c	54	316	15.8	417	2	BF510305	BF510305 UI-H-BI4-
c	55	314.8	15.7	507	9	AQ415290	AQ415290 RPCI-11-1
	56	313.2	15.7	521	9	AQ075172	AQ075172 CIT-HSP-2
c	57	312.8	15.6	806	9	BZ600353	BZ600353 WHACQ31TR
	58	310.4	15.5	697	9	B80940	B80940 CIT-HSP-202
	59	310	15.5	512	9	B59557	B59557 CIT-HSP-342
c	60	308.8	15.4	22715	9	AQ839851	AQ839851 69I15-C79
c	61	308.4	15.4	746	2	BG196213	BG196213 RST15426
c	62	308.2	15.4	631	1	AW859880	AW859880 QV4-CT036
	63	306.6	15.3	627	5	BU198128	BU198128 DCBCPE03
c	64	305.8	15.3	578	7	CR742388	CR742388 CR742388
c	65	305.8	15.3	3439	4	HSM803559	AL832252 Homo sapi
	66	305.6	15.3	675	9	AQ108395	AQ108395 CIT-HSP-2
c	67	305	15.2	816	9	AQ746239	AQ746239 HS_2276_A
	68	304	15.2	711	6	CB962291	CB962291 AGENCOURT
	69	303.8	15.2	690	10	AG091193	AG091193 Pan trogl
	70	303.6	15.2	406	9	B59220	B59220 CIT-HSP-201
	71	303.6	15.2	824	3	BI461111	BI461111 603206993
	72	303.4	15.2	547	9	AQ053359	AQ053359 CIT-HSP-2
	73	303.4	15.2	887	5	BU154572	BU154572 AGENCOURT
	74	303.2	15.2	462	9	AQ076528	AQ076528 CIT-HSP-2
	75	302.6	15.1	669	9	B91834	B91834 CIT-HSP-216
	76	302.6	15.1	956	2	BG687097	BG687097 602638915
	77	301.8	15.1	460	9	B66805	B66805 CIT-HSP-201
	78	301.4	15.1	923	3	BI827335	BI827335 603077816
c	79	300.6	15.0	618	9	AQ676986	AQ676986 HS_2148_B
	80	300.4	15.0	452	9	AQ340370	AQ340370 HS_2226_A
	81	300.2	15.0	599	7	CR545051	CR545051 DKFZp470L
	82	300.2	15.0	891	6	CF596869	CF596869 AGENCOURT
	83	300.2	15.0	961	5	BU168265	BU168265 AGENCOURT
c	84	299.6	15.0	668	3	BQ016167	BQ016167 UI-H-DT1-
	85	299.2	15.0	690	10	AG123732	AG123732 Pan trogl
c	86	298.8	14.9	519	2	BF813114	BF813114 QV4-CI015
	87	298.6	14.9	720	2	BG773308	BG773308 602722076
c	88	298.2	14.9	676	1	AV730039	AV730039 AV730039
	89	297.4	14.9	659	9	BZ749231	BZ749231 207I13-SP
	90	297.2	14.9	815	5	BU568866	BU568866 AGENCOURT
c	91	296.6	14.8	507	7	CN278538	CN278538 170006002
	92	295.4	14.8	730	2	BG181797	BG181797 RST650 At
	93	295.2	14.8	723	5	BX095590	BX095590 BX095590
c	94	295	14.7	704	10	AG171536	AG171536 Pan trogl
	95	294.2	14.7	763	9	AQ371716	AQ371716 RPCI11-15

96	294.2	14.7	915	2	BE729640	BE729640 601562772
97	294	14.7	675	3	BM723230	BM723230 UI-E-EJ0-
98	294	14.7	786	5	BU561203	BU561203 AGENCOURT
99	293.6	14.7	583	9	AQ603128	AQ603128 HS_2118_A
100	293.6	14.7	713	2	BG773035	BG773035 602721322

ALIGNMENTS

RESULT 1
AG115127
LOCUS AG115127 717 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-122E08.F, genomic survey sequence.
ACCESSION AG115127
VERSION AG115127.1 GI:16735646
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 717)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1. .717
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-122E08.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

Query Match 29.6%; Score 592.4; DB 10; Length 717;
Best Local Similarity 97.1%; Pred. No. 1.1e-85;
Matches 635; Conservative 0; Mismatches 16; Indels 3; Gaps 3;

```
Qy        211 GAGCTCAACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGG 270
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         63 GAGCTCAACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCACTGGCCCTTTCAATGG 122

Qy        271 CCTAAAGAGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCC 330
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CCTAAAGAGCTCACCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCC 182

Qy        331 AGCAGGAAGTAGCTACAGCGGTCATCGCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTT 390
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        183 AGCAGGAAGTAGCTACAACGGTCATCGCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTT 242

Qy        391 GGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAA 450
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAA 302

Qy        451 CTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAG 510
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 CTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAG 362

Qy        511 GATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGC 422

Qy        571 GTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCC 630
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 GTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCC 482

Qy        631 CTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCAT 690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 CTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCAT 542

Qy        691 TCTTTGTGTCCACACTA-CCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTC 749
              ||||||||||||||  ||| ||||||||||||||||||||||||||||| ||| ||||
Db        543 TCTTTGTGTCCACCCTACCCTTTATGAGCTGTAACACTCACTGCGAGGGCCTGCGGCTCT 602

Qy        750 ATTCCTGAAGTCAACAGACCACGAA-CCCACTGGAAGGAACAAA-GAACTCCCGATGTGC 807
              ||||||||||||||| ||||||||| |||||||||| ||||||| |||||||||| ||||
Db        603 ATTCCTGAAGTCAACGGACCACGAACCCCACTGGAAAGAACAAAGGAACTCCCGAAGTGC 662

Qy        808 TGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAG 861
                  | ||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 CCGCCTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAG 716
```

RESULT 2
BM041081
LOCUS BM041081 855 bp mRNA linear EST 07-NOV-2001
DEFINITION 603614811F1 NIH_MGC_108 Homo sapiens cDNA clone IMAGE:5556482 5', mRNA sequence.
ACCESSION BM041081
VERSION BM041081.1 GI:16770336
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 855)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2020 row: f column: 03
High quality sequence stop: 819.
FEATURES Location/Qualifiers
source 1. .855
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5556482"
/tissue_type="Wilms' tumor, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_108"
/note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the

laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 25.2%; Score 504.6; DB 3; Length 855;
Best Local Similarity 84.8%; Pred. No. 1.5e-71;
Matches 623; Conservative 0; Mismatches 84; Indels 28; Gaps 4;

```
Qy        334 AGGAAGTAGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGG 392
              ||||||||||||  || |||||||  |||| ||||||||||| |||||||| ||||  |
Db          2 AGGAAGTAGCTAGAGTGGTCATCACCCAATTCCCAACAGCAGTTGGGGTGTCTTGTTAAG 61

Qy        393 AGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACT 452
              ||||  ||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db         62 TGGGGAGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGGGGACTTGGAGAACT 121

Qy        453 TTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGA 512
              |||  ||||||||||  ||||||||||  ||||||||||   || ||||||||||| |||
Db        122 TTTCTGTCTAGCTAGAGGATTGTAAACACACCAATCAGTGCTATGTGTCTAGCTAGAGGT 181

Qy        513 TTGTAAATGC-------------------------ACCAATCAGCACTCTGTAAAATGGA 547
              |||||||||                         |||||||||||||||||||||||||
Db        182 TTGTAAATGCACCAATCAGCACTCTGTAAAAACGGACCAATCAGCACTCTGTAAAATGGA 241

Qy        548 CCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAG 607
              ||||||||  ||||||  ||||  |||  ||||||||||  |||||  |||||||||  |||  |||
Db        242 CCAATCAGTAGGATGCGGGCAGGGCCAAATAAGGGAATAAAAGCTGGCCACCTGAGTCAG 301

Qy        608 CAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCA 667
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||  |||||||
Db        302 CAGTGGCAACCCACTCGGGTCCCCTTCCATGCTGTGGAAGCTTTGTTCTTTCACTCTTCA 361

Qy        668 CAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACT 727
              ||||||||||||||||||||||  ||||||  ||||||||  ||||||||||||||  |||||||
Db        362 CAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACCACCTTTATGAGCTGCAACACT 421

Qy        728 CACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGG 786
              ||||||||  |||||   ||||||  ||||||||||| | ||||||||||||||  || |||
Db        422 CACTGCGAAGGTCTTCAGCTTCACTCCTGAAGTCAGCGAGACCACGAACCCATGGGGAGG 481

Qy        787 AACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCA 846
              |||||   ||| | ||  |||  |||||||||||||||||||||||||||||  ||||
Db        482 AACAATCGACTTCAGACATGCCACCTTTAAGAGCTGTAACACTCACTGCGAAGGTCTGTG 541

Qy        847 GCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACA 906
              ||||||||||||||||||  |||||||  |||||||  ||||||||||  |  |||||||
Db        542 GCTTCACTCCTGAAGTCAGCAAGACCACGAACCCACTGGAAGGAAGAAATTTCGGACACA 601

Qy        907 CCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACC 966
              |||||  |||||||  ||||||||||  |||||  ||||||||  |||  ||||||||||||
Db        602 TCTGAACATCTGAATGAACAAACTCTGGACACGCCATCTTTAAGAACTGTAACACTCACT 661

Qy        967 GCAAGGGTCTGTGGCTTCATTCTTG-AAGTCAGCAAGACCAAGAACCCACCGGAAGGAAC 1025
              |  |||||   |||  ||||||||||  ||||||||||||||||||||||||  |||||||
Db        662 GTGAGGGTTCATGGGTTCATTCTTGAAAGTCAGCAAGACCAAGAACCCACCAGAAGGAAC 721

Qy       1026 AAATTCCAGACACAG 1040
              |||||| |||||||
Db        722 CAATTCCGGACACAG 736
```

RESULT 3
BZ608243/c
LOCUS BZ608243 836 bp DNA linear GSS 08-JUN-2003
DEFINITION WHACT79TR Human MCF7 breast cancer cell line library (MCF7_1) Homo sapiens genomic clone MCF7_1-17N14, genomic survey sequence.
ACCESSION BZ608243
VERSION BZ608243.1 GI:31516804
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 836)
AUTHORS Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q., Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P., Gray,J.W. and Collins,C.
TITLE End-sequence profiling: Sequence-based analysis of aberrant genomes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
PUBMED 12788976
COMMENT Contact: Volik SV
Colin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolik@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .836
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="MCF7_1-17N14"
/sex="female"
/clone_lib="Human MCF7 breast cancer cell line library (MCF7_1)"
/note="Vector: pECBAC1; Site_1: HindIII; This library was constructed from MCF7 breast cancer cell line by Amplicon Express (http://www.genomex.com) using their standard procedure."
ORIGIN

Query Match 23.8%; Score 476.8; DB 9; Length 836;
Best Local Similarity 87.6%; Pred. No. 4.6e-67;
Matches 589; Conservative 0; Mismatches 52; Indels 31; Gaps 5;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        669 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGT 610

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||  ||||||||||||||||||||  ||||||  ||||||||||  ||  ||  | |||||
Db        609 GACATGACTTCTCCCCTTTCTAGGTCCCGTGACATCCATCTTGCTATTACTCACCTTTGG 550

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        549 GCCCTGTATTTTTAACCTCTTTGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 490

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              |||||||  ||||||||||  ||||||||||||||||||||  |||||||  |||||||||
Db        489 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACT---AACTTCTACTGAGGACC 433

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||  | |  ||||||||||  | ||||||||||  ||||||||  ||||||||||
Db        432 CCTGGACCAAACTGCTGGCCCTTTGACTGGCCTAAAGAGTTCCCCTCTAGAGGACACTAC 373

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||||  ||||||||  ||||||||||||||||||||||||  |  |||||||  ||
Db        372 AACTGCAGGGACCCTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAATGGTCATCACCC 313

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              ||  |||||  |||||  |||||||||  |  ||||  |||||  ||  ||||||||||||||||||
Db        312 AATTCCCAGTAGCAGTTGGGGTGTTCCGTTTAGAGGGTGGGTTGAGAGGTGAAGCCAGCT 253

Qy        420 GGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              ||  ||  ||||  |||  ||||||||||||||||||||  ||||||||||||||  ||||||
Db        252 GGATTTCCTGGATCAAGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAAAGGACTGTAAA 193

Qy        479 TGCACCAATCAGCA-----------------------CTCTGTGTCTAGCTAAAGGAT 513
              ||||||||||||||                       ||||||||||||||||||||||
```

```
Db        192 TGCACCAATCAGCACTCTATAAAAATGCACCAATCAGCGCTCTGTGTCTAGCTAAAGGAT 133

Qy        514 TGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTC 573
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db        132 TGTAAACACACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGAC 73

Qy        574 AAATAAGGGAGTAAAAACTGGCCA-CCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCT 632
              || ||||||| ||||| ||||||| ||| |||||||||| ||||||||||||||||||||
Db         72 AAGTAAGGGAATAAAAGCTGGCCACCCCCAGCCAGCAGCGGCAACCCACTCGGGTCCCCT 13

Qy        633 TCCACACTGTGG 644
              ||||||||||||
Db         12 TCCACACTGTGG 1


RESULT 4
AG045844
LOCUS       AG045844                655 bp    DNA     linear   GSS 02-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-024O04.R, genomic survey sequence.
ACCESSION   AG045844
VERSION     AG045844.1  GI:16582736
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 655)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .655
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-024O04.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             22.9%;  Score 458;  DB 10;  Length 655;
  Best Local Similarity   89.2%;  Pred. No. 5.4e-64;
  Matches  505;  Conservative    0;  Mismatches   60;  Indels    1;  Gaps    1;

Qy        489 AGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGAC 548
              ||  |||||| ||||||||||  || ||||| ||||||||||||||||||||||||||
Db         50 AGAGCTCTGTATCTAGCTAAAGATTTATAAATACACCAATCAGCACTCTGTAAAATGGAC 109

Qy        549 CAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGC 608
              ||||||||||||||||  |||| |||||||||| ||||| || |||||| ||||||||
Db        110 CAATCAGCAGGATGTGGATGGGGCCAAATAAGGGAATAAAAGCTAGCCACCTGAGCCAGC 169

Qy        609 AGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCAC 668
              || |||||||||||| |||||||||||| ||||||||||||||||||||||||||||||
Db        170 AGCGGCAACCCACTCAGGTCCCCTTCCACTCTGTGGAAGCTTTGTTCTTTTGCTCTTCAC 229

Qy        669 AATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTC 728
              |||||| |||||||||||||| |||||| ||||| ||||||||||||||||||||||||||
Db        230 AATAAAACTTGCTGCTGCTCACTCTTTGGGTCCATACTACCTTTATGAGCTGTAACACTC 289

Qy        729 ACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGA 787
              ||  |||||||||||||||||||||||||||||| | ||||||| |||||||  | ||||
Db        290 ACCACGAGGGTCTGTGGCTTCATTCCTGAAGTCAGCAAGACCACAAACCCACGAGGAGGA 349

Qy        788 ACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAG 847
              |||||  |||||  |||||||  |||||||||||||||||||||||||| |||| ||||  |
Db        350 ACAAACAACTCTGGATGTGCCACCTTTAAGAGCTGTAACACTCACTGTGAAGGTCTGTGG 409

Qy        848 CTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACAC 907
              |||||||||||||||  |  |||||||  |||||||| ||||||||||||||||||||||
Db        410 CTTCACTCCTGAAGTCGGCAAGACCACGAACCCACCGGAAGGAAGAAACTCTGGACACAT 469

Qy        908 CTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCG 967
              ||||| |||||||||||||||||| |||| |||||||||  |||||  ||||||||||||
Db        470 CTGAACATCTGAAGGAACAAACTCCGGACATACCATCTTTAAGAGCTATAACACTCACCG 529

Qy        968 CAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAA 1027
              ||||||||||  |||||||||||||| |||||||||||||||||||||||||||||||||
Db        530 CAAGGGTCTGCAGCTTCATTCTTGAAATCAGCAAGACCAAGAACCCACCGGAAGGAATAA 589

Qy       1028 ATTCCAGACACAGTAGGAAATCTGTA 1053
              |||||||||||| |  |    ||| |
Db        590 ATTCCAGACACATTTTGGTGACTGCA 615


RESULT 5
CR745511
LOCUS       CR745511                672 bp    mRNA    linear   EST 21-DEC-2004
DEFINITION  CR745511 Soares_testis_NHT Homo sapiens cDNA clone IMAGp998P131862
            ; IMAGE:757500 5', mRNA sequence.
ACCESSION   CR745511
VERSION     CR745511.1  GI:51667764
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   Ebert,L., Heil,O., Hennig,S., Korn,B., Neubert,P., Partsch,E.,
            Peters,M., Radelof,U. and Schneider,D.
  TITLE     I.M.A.G.E. cDNA Clone Collection
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Email: www.rzpd.de
            RZPD; IMAGp998P131862.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
            Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 100
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998P131862 ; IMAGE:757500"
                     /sex="male"
```

```
                     /lab_host="DH10B"
                     /clone_lib="Soares_testis_NHT"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

  Query Match             22.8%;  Score 456.6;  DB 7;  Length 672;
  Best Local Similarity   89.3%;  Pred. No. 8.9e-64;
  Matches  503;  Conservative    0;  Mismatches   59;  Indels    1;  Gaps    1;

Qy        485 AATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAAT 544
              ||||| |||||||||||| | ||||| ||||||||||||||||||||||||||||||||
Db          1 AATCAACACTCTGTGTCTGGATAAAGAATTGTAAATGCACCAATCAGCACTCTGTAAAAT 60

Qy        545 GGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGADC 604
              |||||||||| |||||| ||||||||| ||||||| ||| ||||| |||||||||| |||
Db         61 GGACCAATCATCAGGATATGGGCGGGGACAAATAAAGGAATAAAAGCTGGCCACCCCADC 120

Qy        605 CAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCT 664
              ||||||  ||||||| ||| |||||||||||||||||||||||||||||    ||||
Db        121 CAGCAGCAGCAACCCTCTCAGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTCCACTCT 180

Qy        665 TCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAAC 724
              |||||||||| |||||||||||| |||||| |||  ||||||||||||||||||||||
Db        181 TCACAATAAATATTGCTGCTGCTCACTCTTTGGGTCTGCACTACCTTTATGAGCTGTAAC 240

Qy        725 ACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGA 783
              |||||| |||||||||| |||||||||||||||||||| | ||||||||||||||||||
Db        241 ACTCACCGCGAGGGTCTGCGGCTTCATTCCTGAAGTCAGCGAGACCACGAACCCACTGGG 300

Qy        784 AGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCT 843
              |||||||| |||||  || ||||  ||||||| | |||||||||||||||||||  |||
Db        301 AGGAACAAACAACTCTGGACGTGCCACCTTTAAAAACTGTAACACTCACTGCGAAGGTCT 360

Qy        844 GCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGAC 903
              |  |||||||||||||||||||||||||||  || ||||||||||||||||||||||||||
Db        361 GTGGCTTCACTCCTGAAGTCAGTGAGACCATGAATCCACCAGAAGGAAGAAACTCTGGAC 420

Qy        904 ACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTC 963
              ||| ||||| |||||||||||||||||||| ||| ||||||||| ||||||||||||||||
Db        421 ACATCTGAACATCTGAAGGAACAAACTCCACACATACCATCTTTAAGAGCTGTAACACTC 480

Qy        964 ACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGA 1023
              |||   | |||| | |||||||||||||||||||| ||||||||||||||| |||| ||
Db        481 ACCATGAAGGTCCGCGGCTTCATTCTTGAAGTCAGCGAGACCAAGAACCCACTGGAATGA 540

Qy       1024 ACAAATTCCAGACACAGTAGGAA 1046
              | |||||||||||||  || ||
Db        541 ATAAATTCCAGACACACCAGCAA 563


RESULT 6
AQ745404/c
LOCUS       AQ745404                860 bp    DNA     linear   GSS 16-JUL-1999
DEFINITION  HS_2276_A1_G08_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2276 Col=15 Row=M, genomic survey
            sequence.
ACCESSION   AQ745404
VERSION     AQ745404.1  GI:5522926
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 860)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2276  row: M  column: 15
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 860.
FEATURES             Location/Qualifiers
     source          1. .860
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2276 Col=15 Row=M"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

  Query Match             22.5%;  Score 450.8;  DB 9;  Length 860;
  Best Local Similarity   87.8%;  Pred. No. 7.2e-63;
  Matches  526;  Conservative    0;  Mismatches   68;  Indels    5;  Gaps    3;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||   |||||||| |||||||||||| |||||||||| |||
Db        701 AAAGGCTTCTGAAATCAGAC---GCCTTTCANACTCTTATACCAATCTCTGGAGTTTGGC 645

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||| |||||||||||||||| |||||||||||||| || || | |||||
Db        644 AACATGGCTTCTACCCTTTCTAGGTCCTGTGGCAGCCATCTTGCTATTACTCACCTTTGG 585

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | ||||||||| |||||||||||||| ||||||  |||||
Db        584 GCCCTGTATTTTTAACCTCCTTGTCAAATTTCTTTCCTCTAGGATCAAGGCCAGTAAGCT 525

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||| |||||||||| |||||||||||||| |||||||||
Db        524 ACAGATGGTCTTACAAATGGAACCTCAAATGAGCTTAACTAACAACTTCTACTGAGGACC 465

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||| ||||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db        464 CCTGGACTGACCCACTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 405

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||||||||||| ||| |||||||||||||||||||||| || |||||||  ||
Db        404 AACTGCAGGGCCCCTTCATCATCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATCACCC 345

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||||||||| |||||||||||||| ||||||||||||
Db        344 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGAAAGGTGAAGCCAGCT 285

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              ||||||||  ||| ||||| |||||||||||||||| ||||||||||||||||||||||
Db        284 GGGCTTCTATGTCGGGTGGAGACTTGGAGAACTTTTCTGTCTAGCTAAAGGATTGTAAAT 225

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT 539
              ||| ||||||   |||||||||||||||||||||||||||||||||||||||||||| |||
Db        224 GCATCAATCATTGCTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTATGT 165
```

```
Qy        540 -AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCA 597
              |||| | ||||||||||| ||||     ||  |||  |       | ||||  | |||
Db        164 AAAAACGCACCAATCAGCACTATGTAAAATGGACCAATCAGCACTCTTAAAATGGACCA 106


RESULT 7
AG169322
LOCUS       AG169322                680 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-037P09.TJ, genomic survey
            sequence.
ACCESSION   AG169322
VERSION     AG169322.1  GI:16699000
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 680)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-037P09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN


  Query Match             22.3%;  Score 445.8;  DB 10;  Length 680;
  Best Local Similarity   88.0%;  Pred. No. 4.9e-62;
  Matches  530;  Conservative    0;  Mismatches   68;  Indels    4;  Gaps    4;

Qy        400 ATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGT 459
              | |||| ||||||||||||||| ||||||||  |||||||||||||||||||| |||
Db         80 ACTGAGCGGTGAAGCCAGCTGGGTTTCTGGGTGGGGTGGGGACTTGGAGAACTTTTCTGT 139

Qy        460 CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAÁ 519
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        140 CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGTTTGTAAA 199

Qy        520 TGCACCAATCAGCACTCTGT-AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATA 578
              |||||||||||||||||| |||||||||||||| ||||||||| |||| ||||||
Db        200 CACACCAATCAGCACTCTGTAAAAATGGACCAATCAGTAGGATGTGGGTGGGGCCAAATA 259

Qy        579 AGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACA 638
              ||||| |||||  |||||||||| ||| ||||| | ||||||||| |||| ||||||||
Db        260 AGGGAATAAAAGCTGGCCACCC-AGCTAGCAGGGCCAACCCACTAGGGTACCCTTCCATG 318

Qy        639 CTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTG 698
              |||||||||| |||||||||  ||||||||||||||||    |||||||||| |||||| |
Db        319 CTGTGGAAGCGTTGTTCTTTCACTCTTCACAATAAATCAATCTGCTGCTCACTCTTTGGG 378

Qy        699 TCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAA 758
              ||  |||||||||||||||||||||||||||  ||||||  |||||||||||||| |||||||
Db        379 TCTGCACTACCTTTATGAGCTGTAACACTTTCTGCGAAGGTCTGTGGCTTCACTCCTGAA 438

Qy        759 GTCA-ACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAG 817
              |||| | | ||||||||||||||  | |||||  || ||||||  || |  |  ||||||||
Db        439 GTCAGAGAAACCACGAACCCACCAGGAGGAATGAACAACTCCAGACGCACCACCTTTAAG 498

Qy        818 AGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAA 877
              |||||||||||||||| |||| |||||||||||||||||||||||||  ||||| || |
Db        499 AGCTGTAACACTCACTGTGAAGGTCTGCAGCTTCACTCCTGAAGTCAGCAAGACCGCANA 558

Qy        878 CCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACA 937
              |||||||||||||||||||  |||| | ||||| ||| ||||||||||||||  ||||
Db        559 CCCACCAGAAGGAAGAAACTCCAGACATATCTGAACATCAGAAGGAACAAACTCTGGACA 618

Qy        938 CACCATCTTTCAGAGCTG-TAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTC 996
              |||||||||| ||| ||| |||||||||| | |||||||| ||||||||||||||||||
Db        619 CACCATCTTTAAGAACTGTTAACACTCACCACGAGGGTCTGCGGCTTCATTCTTGAAGTC 678

Qy        997 AG 998
              ||
Db        679 AG 680


RESULT 8
BX507549/c
LOCUS       BX507549                649 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686I1625_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686I1625 5', mRNA sequence.
ACCESSION   BX507549
VERSION     BX507549.1  GI:32040300
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
            Wellenreuther,R., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            s1 sequence also available.
            This clone (DKFZp686I1625) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686I1625"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
```

```
ORIGIN

  Query Match             22.0%;  Score 440.8;  DB 5;  Length 649;
  Best Local Similarity   88.9%;  Pred. No. 3.2e-61;
  Matches  514;  Conservative    0;  Mismatches   52;  Indels   12;  Gaps    3;

Qy        476 AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535
              ||||||||||||||  |||||||||||||||||||| ||||||||||||||||||||||
Db        610 AAATGCACCAATCAGAGCTCTGTGTCTAGCTAAAGGTTTGTAAATGCACCAATCAGCACT 551

Qy        536 CTGT-AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGG 594
              |||| |||||||||||||||||||| |||||| ||| |||||||||| ||||| ||||
Db        550 CTGTAAAAATGGACCAATCAGCAGGACGTGGGCAGGGCCAAATAAGGGAATAAAAGCTGG 491

Qy        595 CCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTT 654
              |||||| |||||||||  |||||||||| ||||||||||||||||||||||||||||||
Db        490 CCACCCAAGCCAGCAGCAGCAACCCACTCAGGTCCCCTTCCACACTGTGGAAGCTTTGTT 431

Qy        655 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTAT 714
              ||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||||||
Db        430 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACTACCTTTAT 371

Qy        715 GAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGA 773
              |||||||||||||||| || |||||||| |||||| |||||||||||| | |||||||||
Db        370 GAGCTGTAACACTCACCGCAAGGGTCTGCAGCTTCGTTCCTGAAGTCAGCAAGACCACGA 311

Qy        774 ACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACT 833
              |||||  | ||||||||| |||||| ||    |  ||||||||||||||||||| ||||
Db        310 GCCCACCAGGAGGAACAAAAAACTCCGGACACACCACCTTTAAGAGCTGTAACACCCACT 251

Qy        834 GCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGA 893
              |||||| |||| |||||||||||||||||||| ||||||||||||||||| |||||||||
Db        250 GCGAAGGTCTGTAGCTTCACTCCTGAAGTCAGCGAGACCACAAACCCACCGGAAGGAAGA 191

Qy        894 AACTCTGGACACACCTGA----------ATATCTGAAGGAACAAACTCCAGACACACCAT 943
              |||||||||||| ||||          | |||||||||||||||||| ||||||||||
Db        190 AACTCTGGACACATCTGAAGGAATGGACACATCTGAAGGAACAAACTCCGGACACACCAT 131

Qy        944 CTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGA 1003
              |||| ||| |||||||||||||||| || |||   ||||||||||||||||||||| |||
Db        130 CTTTAAGAACTGTAACACTCACCGCGAGTGTCCACGGCTTCATTCTTGAAGTCAGCGAGA 71

Qy       1004 CCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGT 1041
              |||||||||||| |||||||| |||||| |||||| |
Db         70 CCAAGAACCCACCAGAAGGAACCAATTCCGGACACATT 33


RESULT 9
B72370
LOCUS       B72370                   621 bp    DNA     linear   GSS 08-APR-1999
DEFINITION  RPCI11-8J21.TP RPCI-11 Homo sapiens genomic clone RPCI-11-8J21,
            genomic survey sequence.
ACCESSION   B72370
VERSION     B72370.1  GI:2711521
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-8J21.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7502924"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-8J21"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             21.5%;  Score 429.8;  DB 9;  Length 621;
  Best Local Similarity   84.9%;  Pred. No. 1.9e-59;
  Matches  527;  Conservative    0;  Mismatches   67;  Indels   27;  Gaps    3;

Qy        291 AGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCG 350
              |||||||||| |||||||||||||||||| |||||||| |||||||||||||||| || |
Db          1 AGGACACTACAACTGCAGGGCCCCTTCTTCGCCCCTATCTAGCAGGAAGTAGCTAGAGTG 60

Qy        351 GTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGT 409
              |||||| ||||| |||||||||||  |||||||||||||| ||||||  ||||||||||
Db         61 GTCATCGGCCAATTCCCAACAGCATTTGGGGTGTCCTGTTTAGAGGGGTTATTGAGAGGT 120

Qy        410 GAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAG 469
              |||||  |||||||||||||||| |||||||||| |||||||||| ||||||||||||
Db        121 GAAGCTGGCTGGGCTTCTGGGTCGGGTGGGGACTTAGAGAACTTTTCTGTCTAGCTAAAG 180

Qy        470 GATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGC------- 522
              |||||||||  ||||||||||  |||||||||||||||||||| |||||||  ||
Db        181 GATTGTAAACACACCAATCAGTGCTCTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATA 240

Qy        523 ------------------ACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTG 564
                                 ||||||  |||||||||||||||||||||||||||  ||||||
Db        241 AGCACTCTGTAAAAATAGACCAATAAGCACTCTGTAAAATGGACCAATCAGCATGATGTG 300

Qy        565 GGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCG 624
              || |||| |||||||||| ||||| ||||||||| |||||||||  ||||||| || |
Db        301 GGTGGGGCCAAATAAGGGAATAAAAGCTGGCCACCTGAGCCAGCAGCAGCAACCCGCTGG 360

Qy        625 GGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCT 684
              ||||| ||||||| |||||  |||||||||||||||||||||  ||||||||||||||||
Db        361 GGTCCTCTTCCACGCTGTGAAAGCTTTGTTCTTTTGCTCTTTGCAATAAATCTTGCTGCT 420

Qy        685 GCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTG 744
              ||||| |||||| | |  |||||||||||||||||||||||||||||| | ||||||
Db        421 GCTCACTCTTTGGGCCTGCACTACCTTTATGAGCTGTAACACTCACTGCAAAGGTCTGCA 480

Qy        745 GCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGAT 803
              |||||  |||||||| || | ||||||| |||||||||| || ||  || |||||  ||
Db        481 GCTTCTCTCCTGAAGCCAGCGAGACCACAAACCCACTGGGAGAAATGAACAACTCTGGAC 540

Qy        804 GTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTC 863
              |||  |||||||||||||||||||||||||||||| |||| ||||||||||||||||||||
Db        541 ATGCCACCTTTAAGAGCTGTAACACTCACTGTGAAGGTCTGCAGCTTCACTCCTGAAGTC 600

Qy        864 AGTGAGACCACAAACCCACCA 884
              ||||||||||  |||||||||
Db        601 AGTGAGACCAGGAACCCACCA 621


RESULT 10
```

```
BQ677568
LOCUS       BQ677568                940 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8498032 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6257721
            5', mRNA sequence.
ACCESSION   BQ677568
VERSION     BQ677568.1  GI:21790247
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 940)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2412  row: h  column: 10
            High quality sequence start: 22
            High quality sequence stop: 675.
FEATURES             Location/Qualifiers
     source          1. .940
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6257721"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_112"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             21.2%;  Score 424.2;  DB 5;  Length 940;
  Best Local Similarity   84.7%;  Pred. No. 1.3e-58;
  Matches  557;  Conservative    0;  Mismatches   69;  Indels   32;  Gaps    6;

Qy        154 TTCCTCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAG 213
              ||  | |  |||||||||| ||||||||||||||||||||||||||| ||||||||||||
Db        263 TTGATTTGAGATCGAGGCTATCAAGCTACAGATGATCTTACAAATGGAACCCCAAATGAG 322

Qy        214 CTCAACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCT 273
              ||||||||||||   || |  |||||||||||||| ||||||||||||||||||||||||
Db        323 CTCAACTAACAA---CTACCAAGGACCCCTGGACCAACCCGCTGGCCCTTTCAATGGCCT 379

Qy        274 AAAGAGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGC 333
              |||||| |||||||||||||||||| ||||||||| || |||||  ||||||||||||
Db        380 AAAGAGTTCCCCTCTGGAGGACACTACAACTGCAGGGTCCTTTCTTTGCCCCTATCCAGC 439

Qy        334 AGGAAGTAGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGG 392
              |||||||||||| || |||||||  |||| ||||||||||| |||||||| ||||  |
Db        440 AGGAAGTAGCTAGAGTGGTCATCACCCAATTCCCAACAGCAGTTGGGGTGTCTTGTTAAG 499

Qy        393 AGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACT 452
              ||||  ||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db        500 TGGGGAGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGGGGACTTGGAGAACT 559

Qy        453 TTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGA 512
               || |||||||||| |||||||||| |||||||||  || ||||||||||| |||
```

```
Db        560 CTTCTGTCTAGCTAGAGGATTGTAAACACACCAATCAGTGCTATGTGTCTAGCTAGAGGT 619

Qy        513 TTGTAAATGC-------------------------ACCAATCAGCACTCTGTAAAATGGA 547
              ||||||||||                         ||||||||||||||||||||||||||
Db        620 TTGTAAATGCACCAATCAGCACTCTGTAAAAACGGACCAATCAGCACTCTGTAAAATGGA 679

Qy        548 CCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAG 607
              |||||||| |||||| |||| ||| |||||||||||| ||||| ||||||||| ||| |||
Db        680 CCAATCAGTAGGATGCGGGCAGGGCCAAATAAGGGAATAAAAGCTGGCCACCTGAGTCAG 739

Qy        608 CAGTGGCAA-CCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTC 666
              ||||||||| |||||||||||||||||||  |||||| ||||||| |||||  ||||||
Db        740 CAGTGGCAACCCCACTCGGGTCCCCTTCCATGCTGTGGGAGCTTTGGTCTTTCACTCTTC 799

Qy        667 ACAATAAATCTTGCTGCTGCT-CATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACA 725
              |||||||||||||||||||| || |||||| |||||||| ||||||||||||||| |||
Db        800 ACAATAAATCTTGCTGCTGCTCCACTCTTTGGGTCCACACCACCTTTATGAGCTGCAACC 859

Qy        726 CTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGG 782
              |||||||||   ||||   |||||| |||||||||| | ||||||| |||||  |||
Db        860 CTCACTGCGGAAGTCTTCAGCTTCACTCCTGAAGTCAGCGAGACCACNAACCCCATGG 917


RESULT 11
HSM805177/c
LOCUS       HSM805177               3739 bp    mRNA    linear   HTC 20-JAN-2005
DEFINITION  Homo sapiens mRNA; cDNA DKFZp434L1127 (from clone DKFZp434L1127).
ACCESSION   AL833881
VERSION     AL833881.1  GI:21739387
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3739)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Weil,B., Amid,C.,
            Osanger,A., Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2005) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp434L1127) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp434L1127
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3739
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434L1127"
                     /tissue_type="testis"
                     /clone_lib="434 (synonym: htes3). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
                     /note="putative transcript"
ORIGIN

  Query Match             21.1%;  Score 421.6;  DB 4;  Length 3739;
  Best Local Similarity   83.8%;  Pred. No. 2.4e-58;
  Matches  560;  Conservative    0;  Mismatches   74;  Indels   34;  Gaps    6;

Qy        401 TTGAGAGGTGAAGCCAGCTGGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGT 459
              |||||||||||||||| |||  ||| |||||||  || ||||||||||||||||||| |||
Db       3241 TTGAGAGGTGAAGCCAACTGAACTTCCTGGGTCCAGTCGGGACTTGGAGAACTTTTCTGT 3182
```

```
Qy        460 CTAGCTAAAGGATTGTAAATG------------------------CACCAATCAGCACT 494
              |||||| ||||||||||||                          ||||||||||||||
Db       3181 CTAGCTGAAGGATTGTAAACACACCAATCAGCACTCTGTAAAAACACACCAATCAGCACT 3122

Qy        495 CTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCA 554
              |||||||||||||||| ||||||| | |||||||||| |||||||||||||||| |||
Db       3121 CTGTGTCTAGCTAAAGGTTTGTAAACGGACCAATCAGCGCTCTGTAAAATGGACCAGTCA 3062

Qy        555 GCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGC 614
              | |||||||||| ||| |||||||||| ||||| |||||||||| |||||| || |||
Db       3061 GTAGGATGTGGGCAGGGCCAAATAAGGGAATAAAAGCTGGCCACCCTAGCCAGTAGCGGC 3002

Qy        615 AACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAA 674
              || |||||||||||||||||||||||||| |||||||||||| |||||||||| ||||
Db       3001 AA-CCACTCGGGTCCCCTTCCACACTGTGGGAGCTTTGTTCTTTCGCTCTTCACAGTAAA 2943

Qy        675 TCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCG 734
              |||   ||||| |   || || |||| |||||||||||||||||||||||||||  |
Db       2942 TCT---TGCTGTGCTCACTCTGGGTCCGCACTACCTTTATGAGCTGTAACACTCACCACA 2886

Qy        735 AGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAG 793
              ||| ||||  ||||||||| ||||||| | |||||||||||||| || |||||||||
Db       2885 AGGATCTGCAGCTTCATTCTTGAAGTCAGCAAGACCACGAACCCACCGGGAGGAACAAAC 2826

Qy        794 AACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCAC 853
              |||||   | ||||  |||||||||||||||||||||||||||  ||||||||||||
Db       2825 AACTCTGAACGTGCCACCTTTAAGAGCTGTAACACTCACTGCGAAAGTCTGCAGCTTCAC 2766

Qy        854 TCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAAT 913
              |||||||||||| |||||||  |||||||||||||||||||||| ||||||| |||||
Db       2765 TCCTGAAGTCAGCGAGACCATGAACCCACCAGAAGGAAGAAACTCCGGACACATCTGAAC 2706

Qy        914 ATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGC---AA 970
              ||   |||||||| |||| |||||||||||||| ||| || ||||||||||||    |
Db       2705 ATTGCAAGGAACACACTCCGGACACACCATCTTTAAGAACTATAACACTCACCGCGAAGA 2646

Qy        971 GGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATT 1030
               |||| ||| |||||||||||||||||| |||||||||||||||||||||||||| ||||
Db       2645 AGGTCCGTGACTTCATTCTTGAAGTCAGCGAGACCAAGAACCCACCGGAAGGAACCAATT 2586

Qy       1031 CCAGACAC 1038
              || |||||
Db       2585 CCGGACAC 2578
```

RESULT 12
CR537266
LOCUS CR537266 577 bp mRNA linear EST 06-JUL-2004
DEFINITION DKFZp459N203_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone DKFZp459N203 5', mRNA sequence.
ACCESSION CR537266
VERSION CR537266.1 GI:49685089
KEYWORDS EST.
SOURCE Pongo pygmaeus (orangutan)
ORGANISM Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE 1 (bases 1 to 577)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
TITLE Pongo pygmaeus mRNA (Wambutt,R., Heubner,D., Mewes,H.W., et al.)
JOURNAL Unpublished (2004)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Agowa (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp459N203) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.
FEATURES Location/Qualifiers
source 1. .577
/organism="Pongo pygmaeus"
/mol_type="mRNA"
/db_xref="taxon:9600"
/clone="DKFZp459N203"
/tissue_type="cortex"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="459 (synonym: pcor1)"
/note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

Query Match 20.8%; Score 415.6; DB 7; Length 577;
Best Local Similarity 87.7%; Pred. No. 3.8e-57;
Matches 465; Conservative 0; Mismatches 64; Indels 1; Gaps 1;

```
Qy        476 AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535
              ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db         48 AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAACACACCAATCAGCACT 107

Qy        536 CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGC 595
              |||||||||||||||||||||||||||||||| | |||||||||| ||||| |||||
Db        108 CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGAGCCAAATAAGGGAATAAAAGCTGGC 167

Qy        596 CACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTC 655
              |||| | | |||||| ||||| |||||| ||||| ||||||||||||||||||||| |||
Db        168 CACCTGCGGCAGCAGCGGCAATCCACTCAGGTCCACTTCCACACTGTGGAAGCTTTATTC 227

Qy        656 TTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATG 715
              |||  |||||||||||||||||||||| |||| |||||| |||  ||||||||||||
Db        228 TTTCACTCTTCACAATAAATCTTGCTGCTCCTCACTCTTTGGGTCTGCACTACCTTTATG 287

Qy        716 AGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAA 774
              |||||||||||||| |||||||||| |||||||| |||||||| | |||||| ||
Db        288 AGCTGTAACACTCACCGCGAGGGTCTGCGGCTTCATTTCTGAAGTCAGCGAGACCACAAA 347

Qy        775 CCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTG 834
              |||||  | ||||| ||| |||||| ||||  |    |||||| ||| ||||||||||
Db        348 CCCACCAGGAGGAAAAAACAACTCCAGATGCACCATGTTTAAGGGCTATAACACTCACTA 407

Qy        835 CGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAA 894
              | ||| ||||| |||||||||| |||||||| ||||||||||||||    |||||||||
Db        408 CAAAGGTCTGCGGCTTCACTCCAGAAGTCAGCGAGACCACAAACCCATGGCAAGGAAGAA 467

Qy        895 ACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCT 954
              | ||  |||||| ||||| |||||||||||||||||  |||||||  ||||| ||||||
Db        468 ATTCCAGACACATCTGAACATCTGAAGGAACAAACTCTGGACACACTGTCTTTAAGAGCT 527

Qy        955 GTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGAC 1004
              |||||||||||| | |||||||| |||||||||||||||||||||| ||||
Db        528 GTAACACTCACCACGAGGGTCTGCGGCTTCATTCTTGAAGTCAGCGAGAC 577
```

RESULT 13
CF597035
LOCUS CF597035 746 bp mRNA linear EST 26-SEP-2003
DEFINITION AGENCOURT_15658367 NICHD_Hs_Ov1 Homo sapiens cDNA clone IMAGE:30704734 5', mRNA sequence.
ACCESSION CF597035
VERSION CF597035.1 GI:36354012
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 746)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

```
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gregory F. Erickson, Ph.D.
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM254  row: d  column: 23
            High quality sequence stop: 548.
FEATURES             Location/Qualifiers
     source          1. .746
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30704734"
                     /tissue_type="Ovary"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NICHD_Hs_Ov1"
                     /note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc);
                     Site_2: SfiI (ggccgcctcggcc); Library is oligo-dT primed
                     and directionally cloned. Granulosa lutein cells aspirated
                     from preovulatory folicles of normal cycling women
                     undergoing ovulation induction for infertility due to male
                     factor and normal doners. The cells were from follicles
                     stimulated with Lupron, FSH and hCG. 5' and 3' adaptors
                     were used in cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 2.23
                     kb (range 1.0-4.5 kb). 14/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA)."
ORIGIN

  Query Match             20.5%;  Score 411;  DB 6;  Length 746;
  Best Local Similarity   85.3%;  Pred. No. 1.9e-56;
  Matches  528;  Conservative    0;  Mismatches   60;  Indels   31;  Gaps    5;

Qy        158 TCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAGCTCA 217
              | |||||||||||| |||||||||||||||||||||||||| ||||||||||||||||
Db        126 TATAGGATCGAGGCTATCAAGCTACAGATGATCTTACAAATGGAACCCCAAATGAGCTCA 185

Qy        218 ACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAG 277
              ||||||||   || |  |||||||||||||| ||||||||||||||||||||||||||
Db        186 ACTAACAA---CTACCAAGGACCCCTGGACCAACCCGCTGGCCCTTTCAATGGCCTAAAG 242

Qy        278 AGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGA 337
              || |||||||||||||||||| ||||||||| || |||||  ||||||||||||||||
Db        243 AGTTCCCCTCTGGAGGACACTACAACTGCAGGGTCCTTTCTTTGCCCCTATCCAGCAGGA 302

Qy        338 AGTAGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGG 396
              |||||||| || |||||||  |||| |||||||||||| |||||||| ||||  | |||
Db        303 AGTAGCTAGAGTGGTCATCACCCAATTCCCAACAGCAGTTGGGGTGTCTTGTTAAGTGGG 362

Qy        397 GGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTG 456
              | |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        363 GAGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTC 422

Qy        457 TGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGT 516
              |||||||||| ||||||||||  ||||||||||  || |||||||||| ||| ||||
Db        423 TGTCTAGCTAGAGGATTGTAAACACACCAATCAGTGCTATGTGTCTAGCTAGAGGTTTGT 482

Qy        517 AAATGCACCAATCAGCACTCTGTAAAA----------------------------TGGACCAA 551
              |||||||||||||||||||||||||||                            ||||||||
Db        483 AAATGCACCAATCAGCACTCTGTAAAAACGGACCAATCAGCACTCTGTAAAATGGACCAA 542

Qy        552 TCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGT 611
              |||| |||||| |||| ||| |||||||||| |||||||||||||| ||| |||||||
Db        543 TCAGTAGGATGCGGGCAGGGCCAAATAAGGGAATAAAAACTGGCCACCTGAGTCAGCAGT 602

Qy        612 GGCAACCCACTCGGGTCCCCTTCCACACTGT-GGAAGCTTTGTTCTTTTGCTCTTCACAA 670
              |||||||||||||||||||||||||  |||| |||| ||||||||||  ||||||||||
Db        603 GGCAACCCACTCGGGTCCCCTTCCATGCTGTGGGAAACTTTGTTCTTTCACTCTTCACAA 662

Qy        671 TAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACAC-TCA 729
              |||||||||||||||||| |||||| ||| ||| |||||| ||||||   |||| |||
Db        663 TAAATCTTGCTGCTGCTCACTCTTTGGGTCCCCACCACCTTTTTGAGCTTGCACACTTCA 722

Qy        730 CTGCGAGGGTCTGTGGCTT 748
              |||||| || ||    |||
Db        723 CTGCGAAGGGCTTCAACTT 741


RESULT 14
AQ390354/c
LOCUS       AQ390354                630 bp    DNA     linear   GSS 06-MAR-1999
DEFINITION  CITBI-E1-2545M11.TR CITBI-E1 Homo sapiens genomic clone 2545M11,
            genomic survey sequence.
ACCESSION   AQ390354
VERSION     AQ390354.1  GI:4361377
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 630)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2545M11.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .630
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2545M11"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
ORIGIN

  Query Match             20.4%;  Score 408.8;  DB 9;  Length 630;
  Best Local Similarity   88.8%;  Pred. No. 4.6e-56;
  Matches  467;  Conservative    0;  Mismatches   52;  Indels    7;  Gaps    2;

Qy        517 AAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAA 576
              ||||| |||||||||||||||||||||||||||||||||||||||||||| ||||| ||||
Db        624 AAATGGACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGACGGGGACAAA 565

Qy        577 TAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA 636
              ||||||| ||||| |||||||| | |||||||||||||||||| ||||||||||||||||
Db        564 TAAGGGAATAAAAGCTGGCCACTCCAGCCAGCAGTGGCAACCCGCTCGGGTCCCCTTCCA 505
```

```
Qy        637 CACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTG 696
              | || || ||||||||||||||| |||||||| |||||||||| ||||||||||||||||
Db        504 CGCTTTGAAAGCTTTGTTCTTTCGCTCTTCATAATAAATCTTGGTGCTGCTCATTCTTTG 445

Qy        697 TGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTG 756
              |||| ||||||||||||||||||||||||||||||| || ||||||||||||||||||
Db        444 GGTCCGCACTACCTTTATGAGCTGTAACACTCACTGCAAGAGTCTGTGGCTTCATTCCTG 385

Qy        757 AAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTA 815
              |||||| | || |||||||||||| || ||||||||| |||||| || |   |  ||||||
Db        384 AAGTCAGCAAGGCCACGAACCCACCGGGAGGAACAAACAACTCCGGACGCACCACCTTTA 325

Qy        816 AGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACA 875
              |||||||||||||||||||| ||| |      |||||||||||||||   |||||||
Db        324 AGAGCTGTAACACTCACTGCAAAGGT------CTTCACTCCTGAAGTCACCAAGACCACG 271

Qy        876 AACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGA 935
              |||||||||||||||||||||||||| ||| |||||  |||||||||||||||||| ||
Db        270 AACCCACCAGAAGGAAGAAACTCTGGATACATCTGAACTTCTGAAGGAACAAACTCCGGA 211

Qy        936 CACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGT 995
              |||||||||||| |||||||||||||||||| |||||||||  |||||||||||||||| ||
Db        210 CACACCATCTTTAAGAGCTGTAACACTCACTGCAAGGGTCCATGGCTTCATTCTTGAGGT 151

Qy        996 CAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGT 1041
              |||| |||||||||||||| || | ||| ||||||| |||||| |
Db        150 CAGCGAGACCAAGAACCCACTGGGAAGAATAAATTCCGGACACATT 105
```

RESULT 15
BQ926428/c

```
LOCUS       BQ926428                899 bp    mRNA    linear   EST 20-AUG-2002
DEFINITION  AGENCOURT_8754346 Lupski_sciatic_nerve Homo sapiens cDNA clone
            IMAGE:6205474 5', mRNA sequence.
ACCESSION   BQ926428
VERSION     BQ926428.1  GI:22341459
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 899)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution:.MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13628  row: g  column: 11
            High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1. .899
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6205474"
                     /sex="male"
                     /tissue_type="sciatic nerve"
                     /dev_stage="adult, 70 yr"
                     /lab_host="DH10B"
                     /clone_lib="Lupski_sciatic_nerve"
                     /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
                     NotI; Site_2: SalI; cDNA made by oligo-dT priming.
                     Directionally cloned using the following adaptors:
                     5'-TCGACCCACGCGTCCG-3' and
                     5'-GACTAGTTCTAGATCGCGAGCGGCCGCCCT(15)-3'. Size selected >
                     1 kb for average insert length 1.87 kb. This is a primary
                     library, non-amplified. Library constructed by Life
                     Technologies and donated  by J. Lupski, M.D./Ph.D. (Baylor
                     College of Medicine) and is available through Life
                     Technologies."
ORIGIN
```

```
  Query Match             20.1%;  Score 401.6;  DB 5;  Length 899;
  Best Local Similarity   87.3%;  Pred. No. 6e-55;
  Matches  440;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        543 ATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGA 602
              |||||||||||||||||||||||    |||| |||||||||| ||||| ||||||||||
Db        661 ATGGACCAATCAGCAGGATGTGCATGGGGCCAAATAAGGGAATAAAAGCTGGCCACCCGA 602

Qy        603 GCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCT 662
              | |||||||||||||  || |||||||||||| |||||||||||||||||||| |||
Db        601 GTCAGCAGTGGCAACCTGCTGGGGTCCCCTTCCACGCTGTGGAAGCTTTGTTCTTTCGCT 542

Qy        663 CTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTA 722
              |||||||||| |||||||||||||||| |||||| |||| |||| ||||||||||||||
Db        541 CTTCACAATAGATCTTGCTGCTGCTCACTCTTTGGGTCCGCACTGCCTTTATGAGCTGTA 482

Qy        723 ACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACTGG 782
              |||||||   |||||||| |||||||||| |||| | ||   ||| ||||||||||
Db        481 ACACTCACCAGGAGGGTCTGCGGCTTCATTCCTCAAGTGAGCAAGACACAAACCCACTGG 422

Qy        783 AAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTC 842
              ||||||||| |||||| || | ||  |||||||||||||||||||||||||||||||||
Db        421 GAGGAACAAACAACTCCAGACGCGCCACCTTTAAGAGCTGTAACACTCACTGCGAAGGTC 362

Qy        843 TGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGA 902
              || |||||||||| ||||| || |||||||| ||||||   |||||||||||||| |||
Db        361 CGCTGCTTCACTCCCGAAGTTAGCGAGACCACGAACCCAATGGAAGGAAGAAACTCCGGA 302

Qy        903 CACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACT 962
              |||| ||||| |||||||  |||||||||||||||||||||||| |||||| ||||| |
Db        301 CACATCTGAACATCTGAACAAACAAACTCCAGACACACCATCTTTAAGAGCTTTAACATT 242

Qy        963 CACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGG 1022
              ||| || |||||||| |||||||||||||||||| |||| |||||||||  ||||||
Db        241 CACTGCGAGGGTCTGCGGCTTCATTCTTGAAGTCAGCGAGACGAAGAACCCAATGGAAGG 182

Qy       1023 AACAAATTCCAGACACAGTAGGAA 1046
              || ||||||||||||||   ||||
Db        181 AATAAATTCCAGACACATCGGGAA 158
```

RESULT 16
BQ678014

```
LOCUS       BQ678014                900 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8033102 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6091116
            5', mRNA sequence.
ACCESSION   BQ678014
VERSION     BQ678014.1  GI:21790693
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 900)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
```

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2333 row: j column: 13
High quality sequence stop: 652.

FEATURES Location/Qualifiers
source 1. .900
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6091116"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 19.9%; Score 397.2; DB 5; Length 900;
Best Local Similarity 84.2%; Pred. No. 3.1e-54;
Matches 507; Conservative 0; Mismatches 63; Indels 32; Gaps 4;

```
Qy        154 TTCCTCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAG 213
              ||  | |  |||||||||| ||||||||||||||||||||||||||| ||||||||||||
Db        252 TTGATTTGAGATCGAGGCTATCAAGCTACAGATGATCTTACAAATGGAACCCCAAATGAG 311

Qy        214 CTCAACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCT 273
              ||||||||||||   || |  ||||||||||||||| ||||||||||||||||||||||
Db        312 CTCAACTAACAA---CTACCAAGGACCCCTGGACCAACCCGCTGGCCCTTTCAATGGCCT 368

Qy        274 AAAGAGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGC 333
              |||||| |||||||||||||||||| |||||||| || |||||  ||||||||||||
Db        369 AAAGAGTTCCCCTCTGGAGGACACTACAACTGCAGGGTCCTTTCTTTGCCCCTATCCAGC 428

Qy        334 AGGAAGTAGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGG 392
              |||||||||||| || |||||||  |||| ||||||||||| |||||||| ||||  |
Db        429 AGGAAGTAGCTAGAGTGGTCATCACCCAATTCCCAACAGCAGTTGGGGTGTCTTGTTAAG 488

Qy        393 AGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACT 452
               |||| |||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db        489 TGGGGAGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGGGGACTTGGAGAACT 548

Qy        453 TTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGA 512
               || |||||||||| ||||||||||  |||||||||  || ||||||||||| |||
Db        549 CTTCTGTCTAGCTAGAGGATTGTAAACACACCAATCAGTGCTATGTGTCTAGCTAGAGGT 608

Qy        513 TTGTAAATGC-------------------------ACCAATCAGCACTCTGTAAAATGG 546
              ||||||||||                         ||||||||||||||||||||||||
Db        609 TTGTAAATGCACCAATCAGCACTCTGTAAAAAACGGACCAATCAGCACTCTGTAAAATGG 668

Qy        547 ACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCA 606
              ||||||||| |||||| |||| ||| | |||||||||| ||||| ||||||||| ||| ||
Db        669 ACCAATCAGTAGGATGCGGGCAGGGGCCAATAAGGGAATAAAAGCTGGCCACCTGAGTCA 728

Qy        607 GCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTC 666
              |||||||||||||||||||||||||| ||   |||||| | ||||||||||   ||||||
Db        729 GCAGTGGCAACCCACTCGGGTCCCCTTTCATGCTGTGGGAACTTTGTTCTTTCACTCTTC 788

Qy        667 ACAATAAATCTTGCTGCTGCTCATTCTTT--GTGTCCACACTACCTTTATGAGCTGTAAC 724
              |||||||||||||||||||||||| || ||  | | || ||| ||||||||||||  ||
Db        789 ACAATAAATCTTGCTGCTGCTCACTCCTTGGGGGCCCCCACCACCTTTATGAGCTGGCAC 848

Qy        725 AC 726
              ||
Db        849 AC 850
```

RESULT 17

BU602065
LOCUS BU602065 816 bp mRNA linear EST 20-SEP-2002
DEFINITION AGENCOURT_10015763 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6496097 5', mRNA sequence.
ACCESSION BU602065
VERSION BU602065.1 GI:23253824
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 816)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2674 row: p column: 18
High quality sequence stop: 635.

FEATURES Location/Qualifiers
source 1. .816
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6496097"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_142"
/note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH_MGC_141). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 19.6%; Score 392.8; DB 5; Length 816;
Best Local Similarity 84.2%; Pred. No. 1.6e-53;
Matches 500; Conservative 0; Mismatches 64; Indels 30; Gaps 4;

```
Qy        161 AGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACT 220
              |||||||||| |||||||||||||||||||||||||| ||||||||||||||||||||
Db        217 AGGATCGAGGCTATCAAGCTACAGATGATCTTACAAATGGAACCCCAAATGAGCTCAACT 276

Qy        221 AACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGC 280
              |||||   || |  |||||||||||||| ||||||||||||||||||||||||||||||
Db        277 AACAA---CTACCAAGGACCCCTGGACCAACCCGCTGGCCCTTTCAATGGCCTAAAGAGT 333

Qy        281 TCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGT 340
              |||||||||||||||||||| ||||||||| || |||||  ||||||||||||||||||
Db        334 TCCCCTCTGGAGGACACTACAACTGCAGGGTCCTTTCTTTGCCCCTATCCAGCAGGAAGT 393

Qy        341 AGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGG 399
              ||||| || |||||||  |||| |||||||||||| ||||||||| ||||  | |||| |
Db        394 AGCTAGAGTGGTCATCACCCAATTCCCAACAGCAGTTGGGGTGTCTTGTTAAGTGGGGAG 453
```

```
Qy        400 ATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGT 459
              ||||||||||||||||||||||||||||||||  |||||||||||||||||||||| |||
Db        454 ATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTCTGT 513

Qy        460 CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAA 519
              ||||||| ||||||||||| ||||||||||  || ||||||||||| ||| |||||||
Db        514 CTAGCTAGAGGATTGTAAACACACCAATCAGTGCTATGTGTCTAGCTAGAGGTTTGTAAA 573

Qy        520 TGC------------------------ACCAATCAGCACTCTGTAAAATGGACCAATCA 554
              |||                        ||||||||||||||||||||||||||||||||
Db        574 TGCACCAATCAGCACTCTGTAAAAACGGACCAATCAGCACTCTGTAAAATGGACCAATCA 633

Qy        555 GCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGC 614
              | |||||| |||| ||| |||||||||  ||||| ||||||||| ||| ||||||||||
Db        634 GTAGGATGCGGGCAGGGCCAAATAAGGGGATAAAAGCTGGCCACCTGAGTCAGCAGTGGC 693

Qy        615 AACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAA 674
              |||||||||||||||||| ||  |||||||||||||| |||||  |||||||||| |
Db        694 AACCCACTCGGGTCCCCTTTCATGCTGTGGAAGCTTTGNTCTTTCACTCTTCACAATNAT 753

Qy        675 TCTTGCTGCTGCTCATTCTTTGTGTCCACACTA-CCTTTATGAGCTGTAACACT 727
                  |      ||||| |||||| |||||||| | |||||||||||| | ||||
Db        754 CTTGCTGCTGGCTCACTCTTTGGGTCCACACCACCCTTTATGAGCTGCACCACT 807
```

RESULT 18
CD685118/c
LOCUS CD685118 635 bp mRNA linear EST 25-JUN-2003
DEFINITION EST1638 human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION CD685118
VERSION CD685118.1 GI:32200775
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 635)
AUTHORS Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and Zeng,Y.-X.
TITLE Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL Unpublished (2003)
COMMENT Contact: YiXin Zeng
Cancer Center
Sun Yat-sen University
651 DongFeng Road East, GuangZhou 510060, China
Tel: 86-1380-9770-743
Fax: 86-20-8775-4506
Email: yxzeng@gzsums.edu.cn.
FEATURES Location/Qualifiers
source 1. .635
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/note="ESTs generated from a normal nasopharynx cDNA library from southern Chinese"
ORIGIN

Query Match 19.5%; Score 390.6; DB 6; Length 635;
Best Local Similarity 86.4%; Pred. No. 3.9e-53;
Matches 443; Conservative 0; Mismatches 69; Indels 1; Gaps 1;

```
Qy        530 AGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAA 589
              ||| |||||||||||||| |||||||||||||||||| |||| |||||||||| |||||
Db        635 AGCTCTCTGTAAAATGGATCAATCAGCAGGATGTGGGTGGGGCCAAATAAGGGAATAAAA 576

Qy        590 ACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCT 649
              | | ||||| |||||||||||||||| | ||| |||||||||||| ||||||||||
Db        575 GCAGTCCACCTGAGCCAGCAGTGGCAACTCGCTCAGGTCCCCTTCCACGCTGTGGAAGCT 516
```

```
Qy        650 TTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACC 709
              |||||||||  |||||||||||||||||||| ||||||| |||||| |||  |||| ||
Db        515 TTGTTCTTTCACTCTTCACAATAAATCTTGCTACTGCTCACTCTTTGGGTCTGCACTGCC 456

Qy        710 TTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGAC 768
              ||| |||||||||||||||| | ||  |||||  |||||| |||||||  || | ||||
Db        455 TTTGTGAGCTGTAACACTCACCGTGAAAGTCTGCAGCTTCACTCCTGAATCCAGCGAGAC 396

Qy        769 CACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACAC 828
              || |||||||  | |||||  || |||||  ||  | | |||||||||||||||||| |
Db        395 CATGAACCCACCAGGAGGAATGAACAACTCTGGACATCCCGCCTTTAAGAGCTGTAACTC 336

Qy        829 TCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAG 888
              |||||| |||| ||||| |||||||||||||||| |||  ||||||  |||||||| |||
Db        335 TCACTGTGAAGGTCTGCGGCTTCACTCCTGAAGCCAGCAAGACCAAGAACCCACCACAAG 276

Qy        889 GAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTC 948
              ||||||||||    ||||| ||||| ||||||||||||||||||||||||||||||||
Db        275 GAAGAAACTCCAAACACATCTGAACATCTGAAGGAACAAACTCCAGACACACCATCTTTA 216

Qy        949 AGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAG 1008
              ||| || |||||||||||| |||||||| | |||||||||||||| ||||||||||||
Db        215 AGAACTTTAACACTCACCACAAGGGTCCGCGGCTTCATTCTTGAAGACAGCAAGACCAAG 156

Qy       1009 AACCCACCGGAAGGAACAAATTCCAGACACAGT 1041
              ||||||||||||||| |||||||||||||| |
Db        155 AACCCACCGGAAGGAACCAATTCCAGACACAAT 123
```

RESULT 19
AL704833/c
LOCUS AL704833 617 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFZp686B2033_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFZp686B2033 5', mRNA sequence.
ACCESSION AL704833
VERSION AL704833.1 GI:19688188
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 617)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (2001)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
sequenced by MediGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No s1 sequence available.
This clone (DKFZp686B2033) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES Location/Qualifiers
source 1. .617
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686B2033"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"
ORIGIN

```
  Query Match             19.5%;  Score 389.4;  DB 1;  Length 617;
  Best Local Similarity   83.6%;  Pred. No. 6.2e-53;
  Matches  486;  Conservative    0;  Mismatches   71;  Indels   24;  Gaps    3;

Qy        398 GGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGT 457
              | | ||||||||||||||||||||| |||||||||| |||||||||||||||||||||| |
Db        578 GTAATGAGAGGTGAAGCCAGCTGGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTCT 519

Qy        458 GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAA--------- 508
              |||||||||||||||||||||||||||||||||| |||||||| ||    |
Db        518 GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCGCTCTGTGTTTAAAGGATTGTAAGCA 459

Qy        509 -----------AGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCA 557
                        |   |  ||||| |||| ||||  |||||||| ||||||||| |||||
Db        458 CACCAATCAGCACTCTGTAAAATGGACCAGTCAGTGCTCTGTAAGATGGACCAACCAGCA 399

Qy        558 GGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAAC 617
              |||||||||| ||| |||||||| || ||||| ||||||||||||||||||||  |||||
Db        398 GGATGTGGGCAGGGCCAAATAAGAGAATAAAAGCTGGCCACCCGAGCCAGCAGCAGCAAC 339

Qy        618 CCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCT 677
              || |  |||||||||||||| ||||| |||||||||||||||||||||||| || ||||
Db        338 CCGCCTGGGTCCCCTTCCACGCTGTGAAAGCTTTGTTCTTTTGCTCTTCACAGTAGATCT 279

Qy        678 TGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGG 737
              ||||||||||| |||| | |||  |||||||||||||||||||||||| |||  | ||
Db        278 TGCTGCTGCTCACTCTTCGGGTCTGCACTACCTTTATGAGCTGTAACACACACCACAAGC 219

Qy        738 GTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAAC 796
              ||| | |||||||||||||||||||| | ||||||| ||||||| || ||||||||| |||
Db        218 GTCCGCGGCTTCATTCCTGAAGTCAGCGAGACCACAAACCCACCGGGAGGAACAAACAAC 159

Qy        797 TCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCC 856
              ||| || ||||  |||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db        158 TCCTGACGTGCCACCTTTAAGAGCTGTAACACTCACTGCGAAGATCTGCGGCTTCACTCC 99

Qy        857 TGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATC 916
              |||||||||||||||||| |||||||| ||||||||    || ||||||| ||||| |||
Db         98 TGAAGTCAGTGAGACCACGAACCCACCGGAAGGAAG---TTCCGGACACATCTGAACATC 42

Qy        917 TGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTA 957
              || |||||||||||  |||||||||||||| |||| ||||
Db         41 TGGAGGAACAAACTCTGGACACACCATCTTTAAGAGTTGTA 1


RESULT 20
AG030015
LOCUS       AG030015                741 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-002E10.F, genomic survey sequence.
ACCESSION   AG030015
VERSION     AG030015.1  GI:16556888
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 741)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .741
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-002E10.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             19.4%;  Score 387.8;  DB 10;  Length 741;
  Best Local Similarity   85.6%;  Pred. No. 1.1e-52;
  Matches  456;  Conservative    0;  Mismatches   72;  Indels    5;  Gaps    2;

Qy        517 AAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAA 576
              ||||| |||||||||| |||||||||||||||| |||||||||| ||||| ||| ||||
Db         85 AAATGGACCAATCAGCGCTCTGTAAAATGGACCTATCAGCAGGATATGGGCAGGGCCAAA 144

Qy        577 TAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA 636
              ||||||| | ||| |||||||||| |||| | || |||||||  |||  ||||||||||
Db        145 TAAGGGAATTAAAGCTGGCCACCCAAGCCGGAAGAGGCAACCTGCTCCAGTCCCCTTCCA 204

Qy        637 CACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTG 696
                |||||||  |||||||| ||| ||||||||||||||||||||||||||||||| ||||||
Db        205 TGCTGTGGACGCTTTGTTATTTAGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTG 264

Qy        697 TGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTG 756
                |||| |||||||||||||||||||||||||||||||| ||||||||  ||||| ||||||
Db        265 GGTCCGCACTACCTTTATGAGCTGTAACACTCACTGCAAGGGTCTGCAGCTTCGTTCCTG 324

Qy        757 AAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTA 815
              |||||| | |||||| |||| |||||  |||||||| |||||| || ||||  |||||
Db        325 AAGTCAGCGAGACCATGAACTCACTGCGAGGAACAAACAACTCCGGACGTGCCACCTTTG 384

Qy        816 AGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACA 875
              || |||||||||||||| |||| ||||  || ||||||||||||||  | |||||
Db        385 AGGGCTGTAACACTCACTGTGAAGGTCTGTGGCGTCACTCCTGAAGTCAGCAATACCACG 444

Qy        876 AACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGA 935
              |||||||  |||||||||||||||||||||||||||| |||| ||||||||||||||||
Db        445 AACCCACTGGAAGGAAGAAACTCTGGACACACCTGAACATCTAAAGGAACAAACTCCAGA 504

Qy        936 CACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGT 995
              |||||||||||| ||||||||||      ||  |||| ||   |||||||||||| ||||
Db        505 CACACCATCTTTAAGAGCTGTAA----TACAACAAGAGTACCTGGCTTCATTCTTTAAGT 560

Qy        996 CAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAAT 1048
              |||| |||||||||||||| ||||||| | |||||  ||||||| || | ||
Db        561 CAGCGAGACCAAGAACCCACTGGAAGGACCCAATTCTGGACACAGAAGCATAT 613


RESULT 21
AG085080
LOCUS       AG085080                650 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-083A08.R, genomic survey sequence.
ACCESSION   AG085080
VERSION     AG085080.1  GI:16636882
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
```

```
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 650)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-083A08.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             19.3%;  Score 386;  DB 10;  Length 650;
  Best Local Similarity   82.5%;  Pred. No. 2.2e-52;
  Matches  520;  Conservative    0;  Mismatches   65;  Indels   45;  Gaps    5;

Qy        211 GAGCTCAACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGG 270
              ||||||||||||||||||| |  |||||||||||||| | | || | |||||||| |||
Db         63 GAGCTCAACTAACAACTTCTAC-CAGGACCCCTGGACCAATCTGCGGTCCCTTTCACTGG 121

Qy        271 CCTAAAGAGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCC 330
              ||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||
Db        122 CCTAAAGAGTTCCCCTCTGGAGGACACTACAACTGCAGGGCCCCTTCTTCACCCCTATCC 181

Qy        331 AGCAGGAAGTAGCTACAGCGGTCATCG-CCAAATCCCAACAGCAGCTGGGGTGTCCTGTT 389
              |||||||||||||| || |||||| | |||| |||||||||||| |||| |||||||||
Db        182 AGCAGGAAGTAGCTAGAGTGGTCATTGCCCAATTCCCAACAGCAGTTGGGTTGTCCTGTT 241

Qy        390 TGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGC-TTCTGGGTCAGGTGGGGACTTGGAG 448
              | |||||||||||||||||||||||||||||| | |||||  ||| ||| ||||||||||
Db        242 TAGAGGGGGGATTGAGAGGTGAAGCCAGCTGGACTTTCTGCATCAAGTGAGGACTTGGAG 301

Qy        449 AACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAA 508
              ||||||                                          ||||||||||||
Db        302 AACTTT----------------------------------------TCTGTCTAGCTAA 320

Qy        509 AGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCG 568
              ||| |||||||  ||||| |||||||||||||||||||||||||| ||||||  |||||.
Db        321 AGGTTTGTAAACCCACCAGTCAGCACTCTGTAAAATGGACCAATCGGCAGGACATGGGCA 380

Qy        569 GGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTC 628
              ||| |||||||||| ||| | ||||||||| |||||| |||||||||||  |||||||
Db        381 GGGCCAAATAAGGGAATAAGAGCTGGCCACCTGAGCCATCAGTGGCAACCCGTTCGGGTC 440

Qy        629 CCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTC 688
              |||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||
Db        441 CCCTTCCACAGTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCTGCTC 500

Qy        689 ATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTT 748
              | |||||| ||   |||||||||||||||||||||||||||  |||||||||||||| |
Db        501 ACTCTTTGGGTTAGCACTACCTTTATGAGCTGTAACACTCACCACGAGGGTCTGTGGTTC 560

Qy        749 CATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGC 807
```

```
              |||||| ||||||| | ||||||| |||||||||    ||||||| |||||| || ||||
Db        561 CATTCCAGAAGTCAGCAAGACCACAAACCCACTGTGGAGAACAAACAACTCCAGACGTGC 620

Qy        808 TGCCTTTAAGAGCTGTAACACTCACTGCGA 837
              ||||||| |||||||||||| |||||||
Db        621 CACCTTTAACAGCTGTAACACTGACTGCGA 650


RESULT 22
CD171662/c
LOCUS       CD171662                846 bp    mRNA    linear   EST 19-MAY-2003
DEFINITION  AGENCOURT_14065683 NIH_MGC_180 Homo sapiens cDNA clone
            IMAGE:30378848 5', mRNA sequence.
ACCESSION   CD171662
VERSION     CD171662.1  GI:30852425
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM445  row: f  column: 09
            High quality sequence stop: 611.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30378848"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;
                     Site_2: EcoRV (destroyed); Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match             18.8%;  Score 376.4;  DB 6;  Length 846;
  Best Local Similarity   83.7%;  Pred. No. 7e-51;
  Matches  492;  Conservative    0;  Mismatches   76;  Indels   20;  Gaps    5;

Qy        431 TCAGGTGGGGACTTGGAGAA-CTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCA 489
              ||  |||| ||  ||||||| ||| |||||||||||  | |||||||||| ||||||||||
Db        698 TCTAGTGGTGAGGTGGAGAACCTTGTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCA 639

Qy        490 GCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACC 549
              |||| ||||                ||| |||||||||| ||||||||||||||||||
Db        638 GCACCCTGTCA----------------AAACAGACCAATCAGCTCTCTGTAAAATGGACC 595

Qy        550 AATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCA 609
              |||||||||||||||| |||| || |||||||| ||||| | |||  ||||||||||||
Db        594 AATCAGCAGGATGTGGGTGGGGCCAGATAAGGGAATAAAAGCAGGCTGCCCGAGCCAGCA 535

Qy        610 GTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACA 669
              |||||||||||||||||||||||| | |||||||||||||||||||||| ||||||  ||
Db        534 GTGGCAACCCACTCGGGTCCCCTTACGCACTGTGGAAGCTTTGTTCTTTCGCTCTTTGCA 475

Qy        670 ATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCA 729
```

```
Db        474 ATAAATCTTGGTGCTGCTCACTCTTCGGGTCCACACTGCCTTTATGAGCTGTAACACTCA 415

Qy        730 CTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAA 788

Db        414 CCCCGAAGGTCTGAAGCTTCACTCCTGAAGCCAGCGAGACCGCGAACCCACTGGGAGGAA 355

Qy        789 CAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGC 848

Db        354 TGAACAACTCCAGATGTGCCGCC-TTAAAAGCTGTAACACTCACTGCGAAGGTCTGCAGC 296

Qy        849 TTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACC 908

Db        295 TTCACTCGTG-AGCCAGCGAGACCACGAACCCACCAGAAGTAAGAAACTCTGAACACATC 237

Qy        909 TGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGC 968

Db        236 CGAACATCAGAAGGAACAAACTCTGGACATGCTGCCTTTAAGAACTGTAATACTTACCGC 177

Qy        969 AAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016

Db        176 AAGGGTCCGCGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACC 129
```

RESULT 23
CD242806
LOCUS CD242806 929 bp mRNA linear EST 22-MAY-2003
DEFINITION AGENCOURT_14126047 NIH_MGC_179 Homo sapiens cDNA clone IMAGE:30383793 5', mRNA sequence.
ACCESSION CD242806
VERSION CD242806.1 GI:31003270
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 929)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM448 row: d column: 10
High quality sequence start: 4
High quality sequence stop: 499.
FEATURES Location/Qualifiers
source 1. .929
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30383793"
/tissue_type="Pituitary"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_179"
/note="Organ: brain; Vector: pCMV-SPORT6.1; Site_1: EcoRV (destroyed); Site_2: NotI; Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.1 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."
ORIGIN

Query Match 18.5%; Score 370; DB 6; Length 929;
Best Local Similarity 84.8%; Pred. No. 7.4e-50;
Matches 451; Conservative 0; Mismatches 75; Indels 6; Gaps 3;

```
Qy        516 TAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAA 575

Db          2 TAAATGGACCATTCAGTGCTCTGTAAAATGGACCAGTCAGCAGGATGTGGGCAGGGCCAA 61

Qy        576 ATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCC 635

Db         62 ATAAGGGAATAAAAGCTGGCCACCTGAGCCAGCAGCCGGAAGC----CGGGTCCGTTCCC 117

Qy        636 ACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTT 695

Db        118 ACATTGTGGAAACTTAGTTCTTTGTCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTT 177

Qy        696 GTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCT 755

Db        178 GGGTCTGCGCTGCCTTTGTGAGCTGTGGCACTCACTGCGAGGGTCTGCGGCTTCACTCCT 237

Qy        756 GAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTT 814

Db        238 GAAGTCAGCGAGACCACAAACCCACTGGGAAGAACAAACAACTTAGGACGTGCCACCTTT 297

Qy        815 AAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTC-AGTGAGACCA 873

Db        298 AAGAGCTGTAACAGTTGCTGCGAAGGTCTGTGGCTTCACTCCTGAAGTCAAGCGAGCCCA 357

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933

Db        358 TGAACCCAATGGAAGGAAGAAACTGTGGACACGTCTGAACATCTGAAGGAACAAACTCCG 417

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993

Db        418 GACACACCATCTTTAAGAACTGTAACACTCACCACGAGGGTCCATGGCTTCGTTCTTGAA 477

Qy        994 GTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045

Db        478 GTCAGAGAGACCAAGAACCCACCGGAAGGAACCAATTCCCGACACACTAGTA 529
```

RESULT 24
BZ600660/c
LOCUS BZ600660 821 bp DNA linear GSS 08-JUN-2003
DEFINITION WHABB74TF Human MCF7 breast cancer cell line library (MCF7_1) Homo sapiens genomic clone MCF7_1-7M4, genomic survey sequence.
ACCESSION BZ600660
VERSION BZ600660.1 GI:31509122
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 821)
AUTHORS Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q., Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P., Gray,J.W. and Collins,C.
TITLE End-sequence profiling: Sequence-based analysis of aberrant genomes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
PUBMED 12788976
COMMENT Contact: Volik SV
Colin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolik@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .821
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```
                     /clone="MCF7_1-7M4"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN

  Query Match             18.5%;  Score 369.6;  DB 9;  Length 821;
  Best Local Similarity   83.0%;  Pred. No. 8.9e-50;
  Matches  460;  Conservative    0;  Mismatches   84;  Indels   10;  Gaps    3;

Qy        453 TTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGT--------CTAG 504
              | |||||||||||||||| |||||||||||||||||||||||||||         | |
Db        657 TCTGTGTCTAGCTAAAGGTTTGTAAATGCACCAATCAGCACTCTGTAAAAACGCACCAAT 598

Qy        505 CTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTG 564
              |   |   |   |||    |||||| ||| ||||||||||||| ||||||||||||||  ||
Db        597 CAGCATTCTGTAAAACAGACCAATAAGCCCTCTGTAAAATGCACCAATCAGCAGGACATG 538

Qy        565 GGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCG 624
              ||| | | |||||||||||| |||||  |||||||| ||||||||||  ||||||  | |
Db        537 GGCAGAGCCAAATAAGGGAATAAAAGGTGGCCACCTGAGCCAGCAG-CGCAACCTGCCCA 479

Qy        625 GGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCT 684
              | ||| ||||||  |||||||| ||||||||| | ||||||||| ||||||||||||||||
Db        478 GATCCACTTCCAGGCTGTGGAAACTTTGTTCTATCGCTCTTCACGATAAATCTTGCTGCT 419

Qy        685 GCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTG 744
              ||||| |||||| |||   |||||||||||||||||||||||||||| || |||||| |
Db        418 GCTCACTCTTTGGGTCTGTACTACCTTTATGAGCTGTAACACTCACTGTGAAGGTCTGCG 359

Qy        745 GCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGAT 803
              |||| | |||||||||||| | |||||||||||||||||| || |||||| |||||| |||
Db        358 GCTTTACTCCTGAAGTCAGCAAGACCACGAACCCACTGGGAGAAACAAACAACTCCGGAT 299

Qy        804 GTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTC 863
              ||||  |||||||||||||| |||||||||||||| ||||  ||||||||||||||||||
Db        298 GTGCCACCTTTAAGAGCTGTGACACTCACTGCGAAGGTCTGTGGCTTCACTCCTGAAGTC 239

Qy        864 AGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGA 923
              || |||||||| |||||||  |||||||||| |||  |||||| |||| ||||||||
Db        238 AGAGAGACCACGAACCCACTGGAAGGAAGAACCTCCAGACACATCTGAACATCTGAAGTC 179

Qy        924 ACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTT 983
              ||||||||  |||||||||||||| ||||||||||||||||  |  |||||| | |||||
Db        178 ACAAACTCTGGACACACCATCTTTAAGAGCTGTAACACTCATTGTGAGGGTCCGCGGCTT 119

Qy        984 CATTCTTGAAGTCA 997
              ||||||||||||||
Db        118 CATTCTTGAAGTCA 105


RESULT 25
BM041043
LOCUS       BM041043                870 bp    mRNA    linear   EST 07-NOV-2001
DEFINITION  603614780F1 NIH_MGC_108 Homo sapiens cDNA clone IMAGE:5556739 5',
            mRNA sequence.
ACCESSION   BM041043
VERSION     BM041043.1  GI:16770310
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 870)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2020  row: p  column: 20
            High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .870
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5556739"
                     /tissue_type="Wilms' tumor, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_108"
                     /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             18.5%;  Score 369.4;  DB 3;  Length 870;
  Best Local Similarity   78.4%;  Pred. No. 9.4e-50;
  Matches  518;  Conservative    0;  Mismatches  111;  Indels   32;  Gaps    5;

Qy        161 AGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACT 220
              |||||||||| |||||||||||||||||||||||| |||||||||||||||||||||||
Db        199 AGGATCGAGGCTATCAAGCTACAGATGATCTTACAAATGGAACCCCAAATGAGCTCAACT 258

Qy        221 AACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGC 280
              |||||   || |  |||||||||||||| |||||||||||||||||||||||||||||||
Db        259 AACAA---CTACCAAGGACCCCTGGACCAACCCGCTGGCCCTTTCAATGGCCTAAAGAGT 315

Qy        281 TCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGT 340
              |||||||||||||||||||| ||||||||| || |||||  ||||||||||||||||||||
Db        316 TCCCCTCTGGAGGACACTACAACTGCAGGGTCCTTTCTTTGCCCCTATCCAGCAGGAAGT 375

Qy        341 AGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGG 399
              ||||| || |||||||  |||| |||||||||||| |||||||||  ||||  | |||| |
Db        376 AGCTAGAGTGGTCATCACCCAATTCCCAACAGCAGTTGGGGTGTCTTGTTAAGTGGGGAG 435

Qy        400 ATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGT 459
              |||||||||||||||||||||||||||||||  |||||||||||||||||||||||| |||
Db        436 ATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTCTGT 495

Qy        460 CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGAT------ 513
              ||||||| ||||||||||  |||||||||  || |||||||||||| ||| |
Db        496 CTAGCTAGAGGATTGTAAACACACCAATCAGTGCTATGTGTCTAGCTAGAGGTTGTAAAT 555

Qy        514 ------------------TGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAG 555
                                |  ||| | ||||||||||||||||||||||||| |||||||
Db        556 GCACCAATCAGCACTCTGTAAAAACGGACCAATCAGCACTCTGTAAAATGGAGCAATCAG 615

Qy        556 CA--GGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGG 613
              ||  |     ||   ||| |||||||||||| ||| ||   | |  | ||| |||||||||
Db        616 TAAGGATGCGGGCAAGGGCCAAATAAGGGAATAACAAGCTGGCCACTGAGTCAGCAGTGG 675

Qy        614 CAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAA 673
              |||||||||||||||| | ||  |||||||||||      ||       ||||||||||
Db        676 CAACCCACTCGGGTCCC--TTCATGCTGTGGAAGCTTGTACTTTCACTCTTTCACAATAA 733

Qy        674 ATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGC 733
              |||||| || || ||| |||||| |  ||||| |||||||||||||| |||| ||||||
Db        734 ATCTTGGTGGTGATCACTCTTTGGGGTCACACCACCTTTATGAGCTGCAACAGTCACTGG 793
```

```
Qy        734 GAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACTGGAAGGAACAAAG 793
              | ||||||  | |   | |   | ||||  ||||||  | ||    ||| |  | |  |
Db        794 AAAGGTCTGCAGTTCACTCCTGAAGGTCAGGAGACCAGAAGCCATGGGGAGGACAAATCG 853

Qy        794 A 794
              |
Db        854 A 854


RESULT 26
BQ680536
LOCUS       BQ680536                937 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8507906 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6294413
            5', mRNA sequence.
ACCESSION   BQ680536
VERSION     BQ680536.1  GI:21793215
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 937)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2499  row: a  column: 06
            High quality sequence stop: 103.
FEATURES             Location/Qualifiers
     source          1. .937
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6294413"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_112"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             18.5%;  Score 369.2;  DB 5;  Length 937;
  Best Local Similarity   82.9%;  Pred. No. 9.9e-50;
  Matches  528;  Conservative    0;  Mismatches   74;  Indels   35;  Gaps    8;

Qy        158 TCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAGCTCA 217
              | |  |||||||||| ||||||||||||||||||||||||||| ||||||||||||||||
Db        274 TTTGAGATCGAGGCTATCAAGCTACAGATGATCTTACAAATGGAACCCCAAATGAGCTCA 333

Qy        218 ACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAG 277
              ||||||||   || |  ||||||||||||||| ||||||||||||||||||||||||||||
Db        334 ACTAACAA---CTACCAAGGACCCCTGGACCAACCCGCTGGCCCTTTCAATGGCCTAAAG 390

Qy        278 AGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGA 337
              || ||||||||||||||||||| |||| |||| || |||||  ||||||||||||||||
Db        391 AGTTCCCCTCTGGAGGACACTACAACTGGAGGGTCCTTTCTTTGCCCCTATCCAGCAGGA 450

Qy        338 AGTAGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGG 396
              |||||||| || |||||||  |||| |||||||||||| |||||||| ||||  | |||
Db        451 AGTAGCTAGAGTGGTCATCACCCAATTCCCAACAGCAGTTGGGGTGTCTTGTTAAGTGGG 510

Qy        397 GGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTG 456
              | |||||||||||||||||||||||||||||||||||  ||||||||||||||||||  ||
Db        511 GAGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTCTTC 570

Qy        457 TGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGT 516
              || ||||||| ||||||| |||  |||||||||  || ||||||||||| ||| ||||
Db        571 TGGCTAGCTAGAGGATTGGAAACACACCAATCAGTGCTATGTGTCTAGCTAGAGGGTTGT 630

Qy        517 AAATGC-------------------------ACCAATCAGCACTCTGTAAAATGGACCAA 551
              ||||||                         |||||||||||||||||||||||||||||||
Db        631 AAATGCACCAATCAGCACTCTGTAAAAACGGACCAATCAGCACTCTGTAAAATGGACCAA 690

Qy        552 TCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGT 611
              |||| |||||| |||| ||| | ||||||||| ||| | ||||||||| ||| ||||||
Db        691 TCAGTAGGATGCGGGCAGGGGCCAATAAGGGA-TAANAGCTGGCCACCTGAGTCAGCAGG 749

Qy        612 GGCAACCCACTCGGGTCCCCTTCCACACTGT-GGAAGCTTTGTTCTTTTGCTCTTCACAA 670
              |||||||||||||||||||| | ||   ||| |||||||||| |||||  |||||||| |
Db        750 GGCAACCCACTCGGGTCCCCCTTCATGGTGTGGGAAGCTTTGGTCTTTCACTCTTCAC-A 808

Qy        671 TAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTA-CCTTTATGAGCTGTAACACTCA 729
              |||||||||||||||||||| |||||| | |||||| | ||||||||||||| |||||||
Db        809 TAAATCTTGCTGCTGCTCACTCTTTGGGGCCACACCACCCTTTATGAGCTGCAACACTCA 868

Qy        730 CTG--CGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC 764
              |||    ||||      |||||| |||||||| || |
Db        869 CTGGCGAAGGGCTCTCAGCTTCACTCCTGAAGGCAGC 905


RESULT 27
BU168299
LOCUS       BU168299                889 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7941328 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011576
            5', mRNA sequence.
ACCESSION   BU168299
VERSION     BU168299.1  GI:22682283
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13201  row: h  column: 09
            High quality sequence stop: 615.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6011576"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_68"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies. "
```

```
ORIGIN

  Query Match           18.3%;  Score 367;  DB 5;  Length 889;
  Best Local Similarity  77.6%;  Pred. No. 2.3e-49;
  Matches  536;  Conservative    0;  Mismatches   90;  Indels   65;  Gaps    5;

Qy        161 AGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACT 220
              ||||||||||||||||||||||||||| |||||||||| |||||||| ||| |||||||
Db         82 AGGATCGAGGCCATCAAGCTACAGATGGTCTTACAAATGGAACCCCAAGTGAACTCAACT 141

Qy        221 AACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGC 280
              |||||||||  |  |||||||||||||| ||||| ||||||||| | ||||||||||||
Db        142 AACAACTTCCACCAAGGACCCCTGGACCAACCCGTTGGCCCTTTGACTGGCCTAAAGAGT 201

Qy        281 TCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGT 340
              |||| |||| |||||||||| | ||||||||||||||||| |||||||||||||| ||||
Db        202 TCCCTTCTGAAGGACACTACAAGTGCAGGGCCCCTTCTTCGCCCCTATCCAGCAGTAAGT 261

Qy        341 AGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGG 399
              |||| || |||||||  |||| |||||||||||| || |||||||||||| || |||||
Db        262 GGCTAGAGAGGTCATCACCCAATTCCCAACAGCAGTTGTGGTGTCCTGTTTAGATGGGGG 321

Qy        400 ATTGAGAGGTGAAGCCAGCTGGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTG 458
              |||||||||||||||||||||||| ||| |||||||| |||||||||||||||||||| ||
Db        322 ATTGAGAGGTGAAGCCAGCTGGACTTCCTGGGTCAAGTGGGGACTTGGAGAACTTTTCTG 381

Qy        459 TCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAA 518
              |||||||| |||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        382 TCTAGCTAGAGGATTGTAAATGCACCAATCAGTGCTCTGTGTCTAGCTAAAGGATTGTAA 441

Qy        519 ATGCACCAATCAGCACTCTGTA------------------------------------ 540
              |||||||||||||||||||||
Db        442 ATGCACCAATCAGCACTCTGTAAAAACGCAGCAATCAGCACTCTGTAACTAGCTAAAGGG 501

Qy        541 ----AAATGGACCAATCAGCAGGATGT------------------GGGCGGGGTCAAATA 578
                  |||||  |||||||||||   |||                  ||  | ||  ||||||
Db        502 TTGTAAATGCACCAATCAGCACTCTGTAAAATGGAACAATCTGCAGGATGTGGGCAAATA 561

Qy        579 AGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCA---ACCCACTCGGGTCCCCTTCC 635
              | ||| |||||  |||||||||| || ||||||  |||   | || ||| | | |||||||
Db        562 ATGGAATAAAAGCTGGCCACCCCAGTCAGCAGCAGCAGGGAACCGCTCAGATACCCTTCC 621

Qy        636 ACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTT 695
              |||| |||||| ||||||||||   ||||| || | |||||||||||||||| |||||
Db        622 ACACCGTGGAAACTTTGTTCTTACCCTCTTGACGAAAAATCTTGCTGCTGCTCACTCTTT 681

Qy        696 GTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCT 755
              | |||||||| ||||||| |||||  |||  ||||  |||  ||||| |||||||||| |
Db        682 GGGTCCACACCACCTTTAAGAGCTACAACGATCACCACGACAGTCTGCGGCTTCATTCTT 741

Qy        756 GAAGTCAACAGACCACGAACCCACTGGAAGG 786
              ||| ||| |    |   || ||||   || |
Db        742 GAAATCAGCGACACCACAAACCACCAAAAAG 772


RESULT 28
AQ356187/c
LOCUS       AQ356187                 555 bp    DNA     linear   GSS 24-JAN-1999
DEFINITION  CITBI-E1-2537D11.TR CITBI-E1 Homo sapiens genomic clone 2537D11,
            genomic survey sequence.
ACCESSION   AQ356187
VERSION     AQ356187.1  GI:4183360
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2537D11"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
ORIGIN

  Query Match           18.2%;  Score 365;  DB 9;  Length 555;
  Best Local Similarity  88.1%;  Pred. No. 5.5e-49;
  Matches  422;  Conservative    0;  Mismatches   50;  Indels    7;  Gaps    2;

Qy        564 GGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTC 623
              || ||||| |||||||||| ||||| |||||||| | |||||||||||||||||| |||
Db        555 GGCCGGGGACAAATAAGGGAATAAAAGCTGGCCACTCCAGCCAGCAGTGGCAACCCGCTC 496

Qy        624 GGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGC 683
              |||||||||||||| || |||||||||||||||| |||||||| |||||||||| |||
Db        495 GGGTCCCCTTCCACGCTTTGGAAGCTTTGTTCTTTCGCTCTTCATAATAAATCTTGGTGC 436

Qy        684 TGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGT 743
              |||||||||||||| |||| |||||||||||||||||||||||||||||||| || ||||||
Db        435 TGCTCATTCTTTGGGTCCGCACTACCTTTATGAGCTGTAACACTCACTGCAAGAGTCTGT 376

Qy        744 GGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGA 802
              ||||||||||||||||| | || |||||||||||| || |||||||| |||||| ||
Db        375 GGCTTCATTCCTGAAGTCAGCAAGGCCACGAACCCACCGGGAGGAACAAACAACTCCGGA 316

Qy        803 TGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGT 862
              |  |  |||||||||||||||||||||||||| ||| |      ||||||||||||||
Db        315 CGCACCACCTTTAAGAGCTGTAACACTCACTGCAAAGGT------CTTCACTCCTGAAGT 262

Qy        863 CAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGG 922
              ||   ||||||| |||||||||||||||||||||||||| ||| |||||  |||||||
Db        261 CACCAAGACCACGAACCCACCAGAAGGAAGAAACTCTGGATACATCTGAACTTCTGAAGG 202

Qy        923 AACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCT 982
              |||||||||  |||||||||||||||| |||||||||||||||  |||||||||  |||||
Db        201 AACAAACTCCGGACACACCATCTTTAAGAGCTGTAACACTCACTGCAAGGGTCCATGGCT 142

Qy        983 TCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGT 1041
              |||||||||| |||||| |||||||||||||| || | ||| ||||||| |||||| |
Db        141 TCATTCTTGAGGTCAGCGAGACCAAGAACCCACTGGGAAGAATAAATTCCGGACACATT 83


RESULT 29
CF785943
LOCUS       CF785943                 789 bp    mRNA    linear   EST 20-OCT-2003
DEFINITION  AGENCOURT_15990004 NIH_MGC_220 Homo sapiens cDNA clone
            IMAGE:30707257 5', mRNA sequence.
ACCESSION   CF785943
VERSION     CF785943.1  GI:37751457
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 789)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Martin, University of Iowa
             cDNA Library Preparation: M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM1071  row: j  column: 02
            High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1. .789
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30707257"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_220"
                     /note="Organ: mixed; Vector: pYX-Asc; Site_1: EcoRI;
                     Site_2: NotI; Library is oligo-dT primed and directionally
                     cloned Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according tomRNA size fraction, ligated with
                     EcoR I adaptor, digested with Not I and then cloned
                     directionally into pYX-Asc vector. Average insert size
                     0.5-1Kb. Adaptors 5'(AATTCGGCACGAGG)3' and 5'd
                     (CCTCGTGCCG)3'. 3' Linker sequence - GCGGCCGCTGAGAGCC T18.
                     Sequencing primers 3'end: T3 promoter primer 5'd
                     (ATTAACCCTCACTAAAGGGA)3'. 5' End: T7 promoter primer 5'd
                     (TAATACGACTCACTATAGGG)3'. Library was constructed in the
                     laboratory of M. Bento Soares. Average insert size 3-4kb
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             18.2%;  Score 365;  DB 6;  Length 789;
  Best Local Similarity   85.8%;  Pred. No. 4.9e-49;
  Matches  452;  Conservative    0;  Mismatches   70;  Indels    5;  Gaps    4;

Qy        517 AAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAA 576
              || || |||||||||| |||||||||||||||||||||| |||||||||||||| |||
Db         42 AAGTGGACCAATCAGCGCTCTGTAAAATGGACCAATCAGCTGGATGTGGGCGGGGCCAAA 101

Qy        577 TAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA 636
              ||||||| ||||| ||| ||||  |||||||||   ||||||| || |||| ||||||||
Db        102 TAAGGGAATAAAAGCTGACCACTGGAGCCAGCAACAGCAACCCGCTGGGGTGCCCTTCCA 161

Qy        637 CACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTG 696
              |   ||||  |||||||| |||   || ||||||||||||||||||| |||||| | ||||
Db        162 CCTCGTGGGGGCTTTGTTGTTTCATTCATCACAATAAATCTTGCTGTTGCTCACTGTTTG 221

Qy        697 TGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGA--GGGTCTGTGGCTTCATTCC 754
              |||||||||||||||||||||||||||||| | ||||  ||||||||||||||||||||
Db        222 GGTCCACACTACCTTTATGAGCTGTAACACTCCCAGCGAGTGGGTCTGTGGCTTCATTCC 281

Qy        755 TGAAGTCAA-CAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813
              |||||||||  || ||| |||||| |||| |||| |||| |||||  |||||||  ||||
Db        282 TGAAGTCAAGGAGCCCATGAACCCTCTGGGAGGAGCAAACAACTCTGGATGTGCCACCTT 341

Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873
              | |||||||||||| ||||||| |||| ||  ||||||||  || |||||||||  |||||||
Db        342 TTAGAGCTGTAACCCTCACTG-GAAGGTCCACAGCTTCATTCTTGAAGTCAGCGAGACCA 400

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGA-CACACCTGAATATCTGAAGGAACAAACTCC 932
              |||||||| |||||||||||||| ||| |||| ||||| ||||||||||||||||||||
Db        401 TGAACCCACCGGAAGGAAGAAACTCCGGATCACATCTGAACATCTGAAGGAACAAACTCT 460

Qy        933 AGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGA 992
              |||||||||||||| |||||||||||||||||| |  | |||||| ||||||||||||||
Db        461 GGACACACCATCTTTAAGAGCTGTAACACTCACTGTGAAGGTCTGCGGCTTCATTCTTGA 520

Qy        993 AGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACA 1039
              ||||||| |||||||||||||  |||||||| ||||||| ||||||
Db        521 AGTCAGCCAGACCAAGAACCCATTGGAAGGAATAAATTCCGGACACA 567


RESULT 30
AQ839854
LOCUS       AQ839854               14771 bp    DNA     linear   GSS 31-MAR-2000
DEFINITION  260L13-C54 CITB Homo sapiens genomic clone 260L13, genomic survey
            sequence.
ACCESSION   AQ839854
VERSION     AQ839854.1  GI:6652486
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 14771)
  AUTHORS   Carpten,J.D., Makalowska,I., Robbins,C.M., Scott,N., Sood,R.,
            Connors,T.D., Bonner,T.I., Smith,J.R., Faruque,M.U., Stephan,D.A.,
            Pinkett,H., Morgenbesser,S.D., Su,K., Graham,C., Gregory,S.G.,
            Williams,H., McDonald,L., Baxevanis,A.D., Klingler,K.W. and
            Landes,G.M.
  TITLE     A 6-Mb high-resolution physical and transcription map encompassing
            the hereditary prostate cancer 1 (HPC1) region
  JOURNAL   Genomics 64 (1), 1-14 (2000)
   PUBMED   10708513
COMMENT     Contact: Carpten JD
            Cancer Genetics Branch
            National Human Genome Research Institute/National Institutes of
            Health
            Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD
            Tel: 301 435 5626
            Fax: 301 435 5465
            Email: jdc@nhgri.nih.gov
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .14771
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="260L13"
                     /clone_lib="CITB"
ORIGIN

  Query Match             18.1%;  Score 363;  DB 9;  Length 14771;
  Best Local Similarity   81.1%;  Pred. No. 4.4e-49;
  Matches  481;  Conservative    0;  Mismatches   85;  Indels   27;  Gaps    4;

Qy        453 TTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGA 512
              | |||||||||||||||||||||||||||||||| |||||||| ||
Db       1665 TCTGTGTCTAGCTAAAGGATTGTAAATGCACCAGTCAGCACTGTGAAA---------- 1712

Qy        513 TTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGT 572
              |||||| |||||||||||||||||||||||||||||||||||||||| ||||||||||| |||
Db       1713 ----AAATGGACCAATCAGCACTCTGTAAAATGGACCAATCAGTAGGATGTGGGCAGGGC 1768

Qy        573 CAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCT 632
              |||||||||| ||||| |||||||||  || |  |||||||| ||  ||| ||||| ||
Db       1769 CAAATAAGGGAATAAAAGCTGGCCACCGAAGGCCCCAGTGGCAGCCTTCTCTGGTCCACT 1828

Qy        633 TCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTC 692
              || |  |||||||||||||||| |||||||| |||||||||||||||||||||||||| ||
```

```
Db       1829 TCTATGCTGTGGAAGCTTTGTTGTTTTGCTCCTCACAATAAATCTTGCTGCTGCTCACTC 1888

Qy        693 TTTGTGTCCAC---ACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTC 749
              |||| |||  |   |||||||||||||||||||||||||||    |||||||| |||||
Db       1889 TTTGGGTCGGCACTACTACCTTTATGAGCTGTAACACTCACCATCAGGGTCTGCAGCTTC 1948

Qy        750 ATTCCTGAAGTCA-ACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCT 808
              |||||||||||||  |||||||  ||||||||  | |||  |||| |||||| || | ||
Db       1949 ATTCCTGAAGTCAGCCAGACCATGAACCCACCAGGAGGGCCAAACAACTCCAGACGCGCC 2008

Qy        809 GCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGA 868
              |||||||||||||||||||||||||||||| ||  |||||||||||||||||||||| ||
Db       2009 ACCTTTAAGAGCTGTAACACTCACTGCGAAGGTCCACAGCTTCACTCCTGAAGTCAGCGA 2068

Qy        869 GACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAA 928
              ||||||  ||| |||| |||||||||||||       |       |  ||||||||||||||
Db       2069 GACCACTAACTCACCGGAAGGAAGAAACT-------CCTAGACACCTCTGAAGGAACAAA 2121

Qy        929 CTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTC 988
              |||  |||||||||||||| |||||||||||||||||  |   | | |||| ||||| |||
Db       2122 TTCCGGACACACCATCTTTAAGAGCTGTAACACTCACTGTGAAGTTCTGCGGCTTCGTTC 2181

Qy        989 TTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGT 1041
              |||||||||||  ||| |||||| |||| |||||| | ||| ||  |||||| |
Db       2182 TTGAAGTCAGCGAGAGCAAGAAACCACTGGAAGGCATAAACTCTGGACACAAT 2234
```

RESULT 31
AQ742559/c

```
LOCUS       AQ742559                764 bp    DNA     linear   GSS 16-JUL-1999
DEFINITION  HS_5384_B1_B11_SP6 RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=960 Col=21 Row=D, genomic survey sequence.
ACCESSION   AQ742559
VERSION     AQ742559.1  GI:5520081
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 764)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 960  row: D  column: 21
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 764.
FEATURES             Location/Qualifiers
     source          1. .764
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=960 Col=21 Row=D"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN
```

```
  Query Match           18.1%;  Score 362.2;  DB 9;  Length 764;
  Best Local Similarity  82.7%;  Pred. No. 1.4e-48;
  Matches  454;  Conservative    0;  Mismatches   83;  Indels   12;  Gaps    3;

Qy        517 AAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAA 576
              ||  | ||||||  |||| ||||||||||||||||||||||||||||||||||| |||| ||||
Db        647 AATGGGACCAATTAGCATTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAA 588

Qy        577 TAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCC----T 632
              ||||||  |||| ||||||  |   | |||||||  |||| ||| | |||| |||
Db        587 TAAGGGGATAAAAGCTGGCCCCAGCAACCAGCAGCAGCAAACCAGTTGGGTTCCCCCTTC 528

Qy        633 TCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTC 692
              | | |  |||||||||||||||||||||||||||||| ||||||||||||||||||  ||
Db        527 CACCCCCCTTGGAAGCTTTGTTCTTTTGCTCTTCACAACAAATCTTGCTGCTGCTCACTC 468

Qy        693 TTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATT 752
              |||  ||||  |||||||||||||||||||||||||||  |||||||||| |||||||||
Db        467 ATTGGGTCCCTACTACCTTTATGAGCTGTAACACTCACCACGAGGGTCTGCGGCTTCATT 408

Qy        753 CCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCC 811
              |||||||||| | |||||||||||||||  | |||||||| |||||| || ||||  ||
Db        407 CCTGAAGTCAGCGAGACCACGAACCCACCAGGAGGAACAAACAACTCCGGACGTGCCACC 348

Qy        812 TTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGAC 871
              |||||||||||||||||||||||| ||| ||||  |||||||||||||||||  |  ||||
Db        347 TTTAAGAGCTGTAACACTCACTGCAAAGGTCTGTGGCTTCACTCCTGAAGTCGGCAAGAC 288

Qy        872 CACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTC 931
              ||  |||||||| ||||||||||||||  ||||||  |||| ||||||||||||||||||
Db        287 CAGGAACCCACCGGAAGGAAGAAACTCCAGACACATTTGAACATCTGAAGGAACAAACTC 228

Qy        932 CAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTG 991
              | ||||||||||||| ||| ||||||         ||  ||||||||||||||||||||
Db        227 CCGACACACCATCTTTAAGAAACTGTAA-------CGTGAGGGTCTGTGGCTTCATTCTTA 175

Qy        992 AAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTG 1051
              |||||||  |||||||||||||||||||||||||  |||||  |||||  || ||   |
Db        174 AAGTCAGTGAGACCAAGAACCCACCGGAAGGAACCAATTCTGGACACGATATGACTACAA 115

Qy       1052 TATTTTTGA 1060
              ||| || ||
Db        114 AATGTTAGA 106
```

RESULT 32
CB984929

```
LOCUS       CB984929                732 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  AGENCOURT_13647005 NIH_MGC_184 Homo sapiens cDNA clone
            IMAGE:30327682 5', mRNA sequence.
ACCESSION   CB984929
VERSION     CB984929.1  GI:30279453
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 732)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
```

```
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: NDCM137  row: b  column: 11
           High quality sequence stop: 488.
FEATURES             Location/Qualifiers
     source          1. .732
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30327682"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_184"
                     /note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site_1:
                     SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc);
                     Library is oligo-dT primed and directionally cloned. cDNA
                     was prepared from a glandular pool of tissues from thyoid,
                     parathyroid, adrenal, cortex and pineal gland.  5' and 3'
                     adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.38
                     kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             17.9%;  Score 357.2;  DB 6;  Length 732;
  Best Local Similarity   80.4%;  Pred. No. 9.1e-48;
  Matches  500;  Conservative    0;  Mismatches   98;  Indels   24;  Gaps    6;

Qy        369 CAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTG 428
              ||  |  |  || |   | || |     |  | ||||||||||||| ||| || |||||||
Db         25 CATTAATTTTGGGGAAATCTTGGTTATTGTTACTGAGAGGTGAAGTCAGATGAGCTTCTG 84

Qy        429 GGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATC 488
              |||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GGTCGGATGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATC 144

Qy        489 AGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGAC 548
              |||||||||           |||||  ||||||||||||||||||| ||||
Db        145 AGCACTCTGT----------------AAAATGGGCCAATCAGCACTCTGTAAAACGGAC 187

Qy        549 CAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGC 608
              ||| ||| |||   ||||| | | |||||||||| ||||| |||||| ||| |||||||
Db        188 CAACCAGTAGGGCATGGGCAGAGCCAAATAAGGGAATAAAAGCTGGCCTCCCAAGCCAGC 247

Qy        609 AGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCAC 668
              ||||||||||  |||||||||||||||||||||| ||||||||||||||||||||||||
Db        248 AGTGGCAACCGGCTCGGGTCCCCTTCCACACTGTGTAAGCTTTGTTCTTTTGCTCTTCAC 307

Qy        669 AATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTC 728
              ||||||||||||||| |||| |||||| |||| ||||||||| | |||||||| |||||
Db        308 CATAAATCTTGCTGCTTCTCACTCTTTGGGTCCGCACTACCTTGAAGAGCTGTACCACTC 367

Qy        729 ACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCACGAACCCACTGGAAGGA 787
              |||||||  ||||||  |||||||  ||||||||||    |||| | ||||||||    ||||
Db        368 ACTGCGAAGGTCTGCGGCTTCACTCCTGAAGTCAGTGAGACAAGGAACCCACCAAGAGGA 427

Qy        788 ACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAG 847
              ||| | ||||||  || | ||  |||||||||||  |||||||||||||||||| ||||| |
Db        428 ACACACAACTCCGGACGCGCCACCTTTAAGAGCCGTAACACTCACTGCGAAGGTCTGCGG 487

Qy        848 CTTCACTCCTGAAGTCA---GTGAGACCACAAACCCACCAGAAGGAAGAAACT-CTGGAC 903
              |||||||||||||||||     ||  |   || |   |  ||||||||||||| ||||||
Db        488 CTTCACTCCTGAAGTCAAGCAAGAACCATGAACCAACCGGGAAGGAAGAAACTCCTGGAC 547

Qy        904 ACACCTGAATATCTGAA-GGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACT 962
              ||| ||||| ||||||| ||||| |||||  |||||||| ||||| | | ||||||||||
Db        548 ACATCTGAACATCTGAAGGGAACCAACTCTGGACACACCTTCTTTTAAAACTGTAACACT 607

Qy        963 CAC-CGCAAGGGTCTGTGGCTT 983
                 | |  |||||||||||||| ||
Db        608 CCCTTGCAAGGGTCTGTGGGTT 629


RESULT 33
AG132879
LOCUS       AG132879                675 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-145G04.F, genomic survey sequence.
ACCESSION   AG132879
VERSION     AG132879.1  GI:16662557
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 675)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .675
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-145G04.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             17.8%;  Score 355.4;  DB 10;  Length 675;
  Best Local Similarity   79.8%;  Pred. No. 1.8e-47;
  Matches  499;  Conservative    0;  Mismatches   81;  Indels   45;  Gaps    5;

Qy        211 GAGCTCAACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGG 270
              |||||||||||||||||| |  |||||||||||||| | | || | |||||||| |||
Db         82 GAGCTCAACTAACAACTTCTACCAAGGACCCCTGGACCAATCTGCGGTCCCTTTCACTGG 141

Qy        271 CCTAAAGAGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCC 330
              ||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||
Db        142 CCTAAAGAGTTCCCCTCTGGAGGACACTACAACTGCAGGGCCCCTTCTTCACCCCTATCC 201

Qy        331 AGCAGGAAGTAGCTACAGCGGTCATCG-CCAAATCCCAACAGCAGCTGGGGTGTCCTGTT 389
              ||||||||||||||| || |||||| | |||| |||||||||||| |||| ||||||||||
Db        202 AGCAGGAAGTAGCTAGAGTGGTCATTGCCCAATTCCCAACAGCAGTTGGGTTGTCCTGTT 261

Qy        390 TGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGC-TTCTGGGTCAGGTGGGGACTTGGAG 448
              | |||||||||||||||||||||||||||||| | |||||  ||   || ||||||||||
Db        262 TAGAGGGGGGATTGAGAGGTGAAGCCAGCTGGACTTTCTGCATCTTCTGAGGACTTGGAG 321

Qy        449 AACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAA 508
```

```
Db        322 AACTTT----------------------------------------TCTGTCTAGCTAA 340

Qy        509 AGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCG 568

Db        341 AGGTTTGTAAACCCACCAGTCAGCACTCTGTAAAATGGACCAATCGGCAGGACATGGGCA 400

Qy        569 GGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTC 628

Db        401 GGGCCAAATAAGGGAATAAGAGCTGGACACCTGAGCCATCAGTGGCCACCCGTTCGGGTC 460

Qy        629 CCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTC 688

Db        461 CCCTTCCACAGTGTGGAAGCTCTGTTCCTTTCATCTTCACAATAAATCTTGCTGCTGCTC 520

Qy        689 ATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTT 748

Db        521 ACTCTTTGCGTTAGCACTACCTTTATGAGCTGTAACA-TCACCAAGACGGTCTGTGGTTC 579

Qy        749 CATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGC 807

Db        580 CATTCCCGAAGTCAGCAACACCCCAAACCCACTTGGGAGAACACAAAATCCCAGACGTGC 639

Qy        808 TGCCTTTAAGAGCTGTAACACTCAC 832

Db        640 CACCTTTACCAGCTGTAACAATGAC 664


RESULT 34
AL603245/c
LOCUS       AL603245                636 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686N032_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686N032 5', mRNA sequence.
ACCESSION   AL603245
VERSION     AL603245.1  GI:15166751
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No s1 sequence available.
            This clone (DKFZp686N032) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686N032"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             17.7%;  Score 354.8;  DB 1;  Length 636;
  Best Local Similarity   86.3%;  Pred. No. 2.3e-47;
  Matches  428;  Conservative    0;  Mismatches   62;  Indels    6;  Gaps    3;

Qy        564 GGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTC 623

Db        636 GGGTGGGGCCAAATAAGGGAAAAAAAGCAGGCCACCCAAGCCAGCAGCGGCAACCTGCTT 577

Qy        624 GGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGC 683

Db        576 GGGTCCCCTTCCACGCTGTGGAAGCTTTGTTCTTTCGCTCTTCACAATAAATCTTGCTGC 517

Qy        684 TGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGT 743

Db        516 TGCTCACTCTTTGGGTCCGCACTACCTTTATGAGCTGTAA----CACTGCGAAGGTCTGC 461

Qy        744 GGCTTCATTCCTGAAGTCA-ACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGA 802

Db        460 AGCTTCACTCCTAAAGCCAGTGAGACCACGAACCCACCAGGAGGAACGAACCACTCCGGA 401

Qy        803 TGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGT 862

Db        400 TGCACCACCTTTAAGAGCTATAACACTCACTGCGAAGGTCTGCGGCTTCACTCCTGAAGT 341

Qy        863 CAGTGAGACCACAAACCCACCAGAA-GGAAGAAACTCTGGACACACCTGAATATCTGAAG 921

Db        340 CAGTGAGACCATGAACCCACCAGAAGGGAAGAAACTCCGGACACATCTGAACATCTGAAG 281

Qy        922 GAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGC 981

Db        280 GAACAAACTCCAGACACACCATCTTTAAGAACTGTAACACTCACCGCGAGAGTCCGCGGC 221

Qy        982 TTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGT 1041

Db        220 TTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCAGAAGGAACCAATTCTGGACATACT 161

Qy       1042 AGGAAATCTGTATTTT 1057

Db        160 AGCATATTGCAACTTT 145


RESULT 35
AQ747612/c
LOCUS       AQ747612                845 bp    DNA     linear   GSS 19-JUL-1999
DEFINITION  HS_5533_A2_C11_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1109 Col=22 Row=E, genomic survey sequence.
ACCESSION   AQ747612
VERSION     AQ747612.1  GI:5534770
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1109  row: E  column: 22
```

```
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 845.
FEATURES              Location/Qualifiers
     source           1. .845
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /clone="Plate=1109 Col=22 Row=E"
                      /sex="male"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             17.4%;  Score 347.2;  DB 9;  Length 845;
  Best Local Similarity   81.1%;  Pred. No. 3.6e-46;
  Matches  428;  Conservative    0;  Mismatches   98;  Indels    2;  Gaps    2;

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              || || || | |   ||   | | |   |   |  ||||| |||||||||  |||||||
Db        530 CAGCACTCTGTAAAATGGACCAATCAGCACTCTGTAAAATGGACCAATCAGTGCTCTGTA 471

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||  | ||||||||||||||| ||||||||| ||||| |||||  |||
Db        470 AAATGGACCAAGAAACAGGATGTGGGCGGGGCCAAATAAGGGAATAAAAGCTGGCTGCCC 411

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
               ||||||||| |||||||  || |||||||||| ||||||||||||||||||||||||||
Db        410 AAGCCAGCAGCGGCAACCACCTGGGGTCCCCTTGCACACTGTGGAAGCTTTGTTCTTTTG 351

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||  |||||||| |||||| |||  ||||||||||||| ||
Db        350 CTCTTCACAATAAATCTTGTGGCTGCTCACTCTTTGGGTCTGCACTACCTTTATGAACTA 291

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCA- 778
              |||||||||||||| ||||||  |||||| |||||||||| | |||||||||| |||
Db        290 TAACACTCACTGCGAAGGTCTGCAGCTTCACTCCTGAAGTCAGCAAGACCACGAATCCAC 231

Qy        779 CTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAA 838
              | || |||||  || | || | |||  |   |||||||||||||||||||| ||| ||
Db        230 CAGGGAGGAATGAACACCTGCAGATGCACCATCTTTAAGAGCTGTAACACTCAATGCAAA 171

Qy        839 GCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTC 898
              | ||||| |||| ||||||||||||||  ||||||  |||||||||||||||||||||||
Db        170 GGTCTGCGGCTTTACTCCTGAAGTCAGCAAGACCAGGAACCCACCAGAAGGAAGAAACTC 111

Qy        899 TGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAA 958
              ||||||  ||||| |||||||||||||||||  ||| |||||||||  ||| ||||||||
Db        110 CAGACACATCTGAACATCTGAAGGAACAAACTCTGGACGCACCATCTTTAAGAACTGTAA 51

Qy        959 CACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCA 1006
              || |||||   ||||||   || ||||||||||| |   |  || ||
Db         50 CAATCACCTTGAGGGTCCTCGGATTCATTCTTGAATTGTTCCGGATCA 3


RESULT 36
BU189030
LOCUS       BU189030                 874 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7966147 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6165027
            5', mRNA sequence.
ACCESSION   BU189030
VERSION     BU189030.1  GI:22703014
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13523  row: b  column: 04
            High quality sequence stop: 685.
FEATURES              Location/Qualifiers
     source           1. .874
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:6165027"
                      /tissue_type="melanotic melanoma"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="NIH_MGC_72"
                      /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 2 kb.   Library constructed by Life
                      Technologies."
ORIGIN

  Query Match             17.3%;  Score 346.4;  DB 5;  Length 874;
  Best Local Similarity   83.7%;  Pred. No. 4.8e-46;
  Matches  462;  Conservative    0;  Mismatches   58;  Indels   32;  Gaps    5;

Qy        154 TTCCTCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAG 213
              ||  | |  ||||||||| ||||||||||||||||||||||||||| ||||||||||||
Db        325 TTGATTTGAGATCGAGGCTATCAAGCTACAGATGATCTTACAAATGGAACCCCAAATGAG 384

Qy        214 CTCAACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCT 273
              ||||||||||||   || |  |||||||||||||| ||||||||||||||||||||||||
Db        385 CTCAACTAACAA---CTACCAAGGACCCCTGGACCAACCCGCTGGCCCTTTCAATGGCCT 441

Qy        274 AAAGAGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGC 333
              ||||||  |||||||||||||||||||| |||||||||  || |||||  ||||||||||
Db        442 AAAGAGTTCCCCTCTGGAGGACACTACAACTGCAGGGTCCTTTCTTTGCCCCTATCCAGC 501

Qy        334 AGGAAGTAGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGG 392
              |||||||||||  || |||||||  |||| ||||||||||||  |||||||||  ||||  |
Db        502 AGGAAGTAGCTAGAGTGGTCATCACCCAATTCCCAACAGCAGTTGGGGTGTCTTGTTAAG 561

Qy        393 AGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACT 452
              ||||  |||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db        562 TGGGGAGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGGGGACTTGGAGAACT 621

Qy        453 TTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGA 512
              ||| ||||||||||  ||||||||||  |||||||||   || ||||||||||||  |||
Db        622 TTTCTGTCTAGCTAGAGGATTGTAAACACACCAATCAGTGCTATGTGTCTAGCTAGAGGT 681

Qy        513 TTGTAAATGC------------------------ACCAATCAGCACTCTGTAAAATGGA 547
              |||||  ||||                        ||||||||||||||||||||||||
Db        682 TTGTANATGCACCAATCAGCACTCTGTAAAAACGGACCAATCAGCACTCTGTAAAATGGA 741

Qy        548 CCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAG 607
              |||||||| ||||||  |||| ||| ||||||||||    ||| ||||||||| ||| |||
Db        742 CCAATCAGTAGGATGCGGGCAGGGCCAAATAAGGGAAATAAAGCTGGCCACCTGAGTCAG 801

Qy        608 CAGTGGCAACCCACTCGGG--TCCCCTTCCACACTGTGGAAG-CTTTGTTCTTTTGCTCT 664
              ||||||||||||||||||    ||| |||||  |||||||| |||||  |||||  ||||
Db        802 CAGTGGCAACCCACTCGGGGTCCCCTTTCCATGCTGTGGAAGCCTTTGNTCTTTCACTCT 861

Qy        665 TCACAATAAATC 676
              |  |||||||||
Db        862 TTCCAATAAATC 873
```

```
RESULT 37
AQ587990/c
LOCUS       AQ587990               607 bp    DNA     linear   GSS 07-JUN-1999
DEFINITION  CITBI-E1-2643I7.TF CITBI-E1 Homo sapiens genomic clone 2643I7,
            genomic survey sequence.
ACCESSION   AQ587990
VERSION     AQ587990.1  GI:5014670
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2643I7.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2643I7"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
ORIGIN

  Query Match             17.3%;  Score 345.8;  DB 9;  Length 607;
  Best Local Similarity   81.8%;  Pred. No. 6.6e-46;
  Matches  423;  Conservative    0;  Mismatches   92;  Indels    2;  Gaps    2;

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              || || || | |   ||   | | |   |   |  ||||| |||||||||  |||||||
Db        517 CAGCACTCTGTAAAATGGACCAATCAGCACTCTGTAAAATGGACCAATCAGTGCTCTGTA 458

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              |||||||||||  | ||||||||||||||| |||||||||| ||||| |||||  |||
Db        457 AAATGGACCAAGAAACAGGATGTGGGCGGGGCCAAATAAGGGAATAAAAGCTGGCTGCCC 398

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              |||||||||  |||||||  || ||||||||||  ||| |||||||||||||||||||||||
Db        397 AAGCCAGCAGCGGCAACCACCTGGGGTCCCCTTGCACGCTGTGGAAGCTTTGTTCTTTTG 338

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||  |||||||| |||||| |||  |||||||||||||| ||
Db        337 CTCTTCACAATAAATCTTGTGGCTGCTCACTCTTTGGGTCTGCACTACCTTTATGAACTA 278

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCA- 778
              |||||||||||||| ||||||  |||||| |||||||||| | ||||||||| |||
Db        277 TAACACTCACTGCGAAGGTCTGCAGCTTCACTCCTGAAGTCAGCGAGACCACGAATCCAC 218

Qy        779 CTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAA 838
              | || |||||  || | || | ||||  |  |||||||||||||||||||| ||| ||
Db        217 CAGGGAGGAATGAACACCTGCAGATGCACCATCTTTAAGAGCTGTAACACTCAATGCAAA 158

Qy        839 GCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTC 898
              | ||||| |||| |||||||||||||  ||||||  |||||||||||||||||||||||
Db        157 GGTCTGCGGCTTTACTCCTGAAGTCAGCAAGACCAGGAACCCACCAGAAGGAAGAAACTC 98

Qy        899 TGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAA 958
              | |||||| ||||| |||||||||||||||||||   || |||||||||| ||| ||||||
Db         97 TAGACACATCTGAACATCTGAAGGAACAAACTCTGAACGCACCATCTTTAAGAACTGTAA 38

Qy        959 CACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGT 995
              || |||||   ||||||   |||||||||||||| |
Db         37 CAATCACCTTGAGGGTCCTCGGCTTCATTCTTGAATT 1


RESULT 38
AQ239957/c
LOCUS       AQ239957               680 bp    DNA     linear   GSS 30-SEP-1998
DEFINITION  CIT-HSP-2387K11.TF.1 CIT-HSP Homo sapiens genomic clone 2387K11,
            genomic survey sequence.
ACCESSION   AQ239957
VERSION     AQ239957.1  GI:3672155
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2387K11.TR.1
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2387K11"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             17.1%;  Score 341.4;  DB 9;  Length 680;
  Best Local Similarity   84.0%;  Pred. No. 3.3e-45;
  Matches  421;  Conservative    0;  Mismatches   76;  Indels    4;  Gaps    3;

Qy        517 AAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAA 576
              |||   ||||||  || ||||||||||||||||||||||||||||| |||| |||| || |
Db        629 AAACAGACCAATGGGCTCTCTGTAAAATGGACCAATCAGCAGGATTTGGGTGGGGCCAGA 570

Qy        577 TAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA 636
              |||| || ||||| | |||  || |||||||||||||||||| || |||| |||||| |
Db        569 TAAGAGAATAAAAGCAGGCTGCCAGAGCCAGCAGTGGCAACCCTCTTGGGTTCCCTTCGA 510

Qy        637 CACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTG 696
```

```
Db        509 CACTGTGGAAGCTTTGTTCTTTTGCTCTTTACAATAAATCTTCTTGCTGCTCACTCTTTG 450

Qy        697 TGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTG 756

Db        449 GGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAAGGTCTGCAGCTTCACTCCTG 390

Qy        757 AAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTA 815

Db        389 AAGCCAGCAAGACCATGAACCCATCAGGAGGAATGAACAACTCCAGATGTGC--ACTTTA 332

Qy        816 AGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACA 875

Db        331 AGAGCTGTAACACTCACCGCAAAGGTCTGCAGCTTCAATCCTG-AGCCAGCGAGACCACG 273

Qy        876 AACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGA 935

Db        272 AACCCACCAGAAAGAAGAAACTCCAAACACATCCGAACATCAGAAGGAACAAGCTCCGGA 213

Qy        936 CACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGT 995

Db        212 CATGCTGCCTTAAAGAACTGTAACACTCACTGCGAGGGTCCGCGGCTTCATTCTTAAAGT 153

Qy        996 CAGCAAGACCAAGAACCCACC 1016

Db        152 CAGTGAGACCAAGAACTCAGC 132
```

```
RESULT 39
B66172/c
LOCUS       B66172                  566 bp    DNA     linear   GSS 21-JUN-1998
DEFINITION  CIT-HSP-2024E20.TR CIT-HSP Homo sapiens genomic clone 2024E20,
            genomic survey sequence.
ACCESSION   B66172
VERSION     B66172.1  GI:2640150
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT-HSP-2024E20.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .566
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7046518"
                     /db_xref="taxon:9606"
                     /clone="2024E20"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
```

ORIGIN

```
  Query Match             17.0%;  Score 340.6;  DB 9;  Length 566;
  Best Local Similarity   81.1%;  Pred. No. 4.7e-45;
  Matches  455;  Conservative    0;  Mismatches   54;  Indels   52;  Gaps    3;

Qy        141 TGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGT 200

Db        565 TTGTCAAATTTGTTTCCTCTAAAATCAAGGCAATCAAGCTACAGATGGTCTTACAAATGG 506

Qy        201 AACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCC 260

Db        505 AACCCCGAATGAGTTCAACTAACAACTTCTACTGAGGACCCCTGGACCGACCCGCTGGCC 446

Qy        261 CTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTC 320

Db        445 ATTCCACTGGCCTAAAGAGTTTCCCTCTGGAGGACACCACAACTGCAGGGCCCCTGCTTC 386

Qy        321 ACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG--CCAAATCCCAACAGCAGCTGG 378

Db        385 ACCCCTATCCAGCAGGAAGTAGCTAGAGCAGTCATTGGCCAAATTCCAAACAGCAGTTGG 326

Qy        379 GGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGG 438

Db        325 GGTGTCCTGTTTAGAGGGGAAATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGG 266

Qy        439 GGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT 498

Db        265 GGACTTGGAGAACTTTTCTGTCTAGCTAAAGGATTGTAAATGCACCAGTCAGTGCTCTGT 206

Qy        499 GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCA---------------------------- 533

Db        205 GTCTAGCTAAAGGTTTGTAAATGCACCAATCAGCACTCTATAAAAATGGACCAATCAGCA 146

Qy        534 ------------------------CTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGG 569

Db        145 CTCTGTAAAATGGACCAATCAGCTCTCTGCAAAATGGATCACTCAGTAGGATGCAGGTGG 86

Qy        570 GGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCC 629

Db         85 GGCCAAATAAGGGAATAAAAGCAGGCCACCTGCACCACCAGCAGCAACCCACTTGGGTCC 26

Qy        630 CCTTCCACACTGTGGAAGCTT 650

Db         25 CC-TCCACACAGTGGAAGCTT 6
```

```
RESULT 40
AQ481794/c
LOCUS       AQ481794                497 bp    DNA     linear   GSS 24-APR-1999
DEFINITION  RPCI-11-237L13.TJ RPCI-11 Homo sapiens genomic clone
            RPCI-11-237L13, genomic survey sequence.
ACCESSION   AQ481794
VERSION     AQ481794.1  GI:4669198
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-237L13.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
```

```
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7590900"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-237L13"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             17.0%;  Score 340.4;  DB 9;  Length 497;
  Best Local Similarity   83.2%;  Pred. No. 5.2e-45;
  Matches  411;  Conservative    0;  Mismatches   81;  Indels    2;  Gaps    2;

Qy        523 ACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGG 582
              |||||||||| ||||||||||||||||||||||||||||||||||| |||| || |||| ||
Db        493 ACCAATCAGCTCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAGATAAAGG 434

Qy        583 AGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGT 642
              ||||||| | |||  ||| |||||||| ||||||||  |||||||||  |||||  |||| 
Db        433 AGTAAAAGCAGGCTGCCCCAGCCAGCACTGGCAACCTTCTCGGGTCCGGTTCCATGCTGT 374

Qy        643 GGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCA 702
              |||||  ||| ||| ||  || ||  ||||||||||||||||||||| ||||||  || || 
Db        373 GGAAGGTTTATTCCTTCACTTTTTGCAATAAATCTTGCTGCTGCTCACTCTTTGGGTTCA 314

Qy        703 CACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA 762
              | || |||||||||||||||||| ||||||| |||||   |||||  |||||| | ||
Db        313 CGCTGCCTTTATGAGCTGTAACACACACTGCGAAGGTCTACAGCTTCTCTCCTGAGGCCA 254

Qy        763 AC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCT 821
               | |||||||||||||||||| |||||  || |||| | ||    ||||| |||||||||
Db        253 GCGAGACCACGAACCCACTGGGAGGAATGAACAACTGCAGACACACTGCC-TTAAGAGCT 195

Qy        822 GTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCA 881
              |||||||||  |||||| |||||||||||||||||| || |||  ||||||  ||||||
Db        194 GTAACACTCAACGCGAAGGTCTGCAGCTTCACTCCTGTAGCCAGAAAGACCATGAACCCA 135

Qy        882 CCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACC 941
              |||||||||||||||||| ||||| | ||| ||| |||||||||||||| ||||  |
Db        134 CCAGAAGGAAGAAACTCTGAACACATCCGAACATCAGAAGGAACAAACTCCGGACATGTC 75

Qy        942 ATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAA 1001
                |||| ||| |||||||| |||| || || ||| | ||| ||||||||||||||||  |
Db         74 GCCTTTAAGAACTGTAACAGTCACTGCGAGCGTCCGCGGCCTCATTCTTGAAGTCAGTGA 15

Qy       1002 GACCAAGAACCCAC 1015
              |||||||||||||
Db         14 GACCAAGAACCCAC 1


RESULT 41
BE393102
LOCUS       BE393102                 666 bp    mRNA    linear   EST 21-JUL-2000
DEFINITION  601309151F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3630709 5',
            mRNA sequence.
ACCESSION   BE393102
VERSION     BE393102.1  GI:9338467
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM319  row: e  column: 14
            High quality sequence stop: 609.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3630709"
                     /tissue_type="endometrium, adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_44"
                     /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies)."
ORIGIN

  Query Match             16.9%;  Score 338.4;  DB 2;  Length 666;
  Best Local Similarity   83.9%;  Pred. No. 1e-44;
  Matches  418;  Conservative    0;  Mismatches   76;  Indels    4;  Gaps    3;

Qy        520 TGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAA 579
              || |||| || || |||||||||||||||||||||||||| |||| |||| || ||||
Db          1 TGGACCAGTCGGCTCTCTGTAAAATGGACCAATCAGCAGGATATGGGTGGGGCCAGATAA 60

Qy        580 GGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACAC 639
              | || ||||| | |||  || ||    |||||||||||| |||||||| ||||| |||
Db         61 GAGAATAAAAGCAGGCTGCCGGATAACGCAGTGGCAACCCGCTCGGGTCCACTTCCGCAC 120

Qy        640 TGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGT 699
              |||||||  |||||||||||  || ||  | ||||||||||  |||||| |||||| ||
Db        121 TGTGGAAACTTTGTTCTTTCTCTGTTTGCCATAAATCTTGCCACTGCTCACTCTTTGGGT 180

Qy        700 CCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAG 759
              ||||||| |||||||||| |||||||||||| |||| ||||||  |||||| |||||||
Db        181 CCACACTGCCTTTATGAGTTGTAACACTCACCGCGAAGGTCTGCAGCTTCACTCCTGAAG 240

Qy        760 TCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGA 818
              || | |||||||||||||||| |||||||| |||||| || |  | ||| ||||||
Db        241 CCATCGAGACCACGAACCCACTGGGAGGAACAAACAACTCCAGAGGCTCCGCC-TTAAGA 299

Qy        819 GCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAAC 878
              |||||||||||||| ||||||  |||||||||||||||||  ||   ||||||||||||
Db        300 GCTGTAACACTCACCGCGAAGGCCTGCAGCTTCACTCCTGA--TCCAGGAGACCACAAAC 357

Qy        879 CCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACAC 938
              ||||| |||||||||||||   ||||| | ||| ||| |||||||||||||||| |||||
Db        358 CCACCGGAAGGAAGAAACTCCTAACACATCCGAACATCAGAAGGAACAAACTCCGGACAC 417

Qy        939 ACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAG 998
              |||  |||| ||| || |||||||||| | |||||| | |||||||||||||| ||||||
Db        418 ACCGCCTTTAAGAACTATAACACTCACCTCGAGGGTCCGCGGCTTCATTCTTGCAGTCAG 477
```

```
Qy        999 CAAGACCAAGAACCCACC 1016
              ||||||||||||||||||
Db        478 TGAGACCAAGAACCCACC 495


RESULT 42
AI276815/c
LOCUS       AI276815                424 bp    mRNA    linear   EST 28-JAN-1999
DEFINITION  qk61d07.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1873453 3'
            similar to contains PTR5.t2 PTR5 repetitive element ;, mRNA
            sequence.
ACCESSION   AI276815
VERSION     AI276815.1  GI:3899089
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 873   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 406.
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1873453"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Co8"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     colon adenocarcinoma, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo. "
ORIGIN

  Query Match             16.9%;  Score 338.2;  DB 1;  Length 424;
  Best Local Similarity   90.2%;  Pred. No. 1.2e-44;
  Matches  385;  Conservative    0;  Mismatches   38;  Indels    4;  Gaps    2;

Qy         81 TAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGGGCCCTGTATTTTTAACCTCT 140
              |||||||  || ||||||||||||||| || |||| ||  ||||||||||||||||||
Db        424 TAGGTCCCATGGCAGCCATCTTGCTATTACTCGCCTTCAGGCCCTGTATTTTTAACCTT- 366

Qy        141 TGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGT 200
              |||||||||||||||||||||||||| ||||||||||||||||| ||||||||||
Db        365 --GTCAAATTTGTTTCCTCTAGGATCGAAGCCATCAAGCTACAGATGGTCTTACAAATGG 308

Qy        201 AACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCC 260
              |||||||||||||||||| |||||||||| |||||||||||||||| |||| ||||||
Db        307 AACCCCAAATGAGCTCAAATAACAACTTCTACTGAGGACCCCTGGACCTACCCACTGGCC 248

Qy        261 CTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTC 320
              |||||| |||||||||||| |||||||  |||||||||| |||||| ||||||||||||
Db        247 CTTTCACTGGCCTAAAGAGTTCCCCTCCCGAGGACACTACAACTGCAAGGCCCCTTCTTC 188

Qy        321 ACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CCAAATCCCAACAGCAGCTGGG 379
              |||| |||||||||||||||| |||||| |||| | |||| |||||||||||| ||||
Db        187 GCCCCCATCCAGCAGGAAGTAGCCACAGCGATCATTGCCCAATTCCCAACAGCAGTTGGG 128

Qy        380 GTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGG 439
              |||||||||||| ||||||| |||||||||||||||||||||||||||||||||| ||||||
Db        127 GTGTCCTGTTTAGAGGGGGCATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCGGGTGGG 68

Qy        440 GACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTG 499
              ||||| |||||||||| |||||||||||||||||||||  ||||||| ||||||||||
Db         67 GACTTAGAGAACTTTTCTGTCTAGCTAAAGGATTGTAAACACACCAATGAGCACTCTGTG 8

Qy        500 TCTAGCT 506
              ||||| |
Db          7 TCTAGTT 1


RESULT 43
AG113408/c
LOCUS       AG113408                681 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-120A12.R, genomic survey sequence.
ACCESSION   AG113408
VERSION     AG113408.1  GI:16733927
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 681)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-120A12.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             16.9%;  Score 337.4;  DB 10;  Length 681;
  Best Local Similarity   78.3%;  Pred. No. 1.5e-44;
  Matches  474;  Conservative    0;  Mismatches   86;  Indels   45;  Gaps    4;

Qy        452 TTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGT--------CTA 503
              || |||||||||||||| ||||||||| ||||||||| |||||||          | |
Db        649 TTCTGTGTCTAGCTAAAGGTTTGTAAATGTACCAATCAGTACTCTGTAAAAATGGACCAA 590
```

```
Qy        504 GCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGT 563
Db        589 TCAGCACTCTGTAAAACGGACCAATCAGCACTCTGTAAAAGGGACTAATCAGCAGGATGT 530

Qy        564 GGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTC 623
Db        529 GAGTGGGGCCAAATAAGGGAATAAAAGCAGGCCACCCAAGCCAGCAGCAGCAACCCTCTT 470

Qy        624 GGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGC 683
Db        469 GGGTCCCCTTCCACGCTGTGGAAGCTTTGTTGTTTTGCTCTTTAGAATAAATCTTGCTGC 410

Qy        684 TGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGT 743
Db        409 TGCTCACTAGTTAAGTCCACACTAACTTTATGTGCTGTAACACTCACCACGAAGGTCTGC 350

Qy        744 GGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGA 802
Db        349 AGCTTCACTCCTGAAGTCAGCAAAAGTATGAACCCACCAGGAGGAATGAACGACTCCGGA 290

Qy        803 TGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGT 862
Db        289 CCTGCCACCTTTAAGAGCTGTAACACTCACTGTGAAGGTCTGCAGCTTAACTCCTGAAGT 230

Qy        863 CAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGG 922
Db        229 CAGCAAGACTACAAAC----------------------------------ATCTGAAGG 205

Qy        923 AACAAACTCCAGAC-ACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGC 981
Db        204 AACAAACTCTGGACGACACCATCTTTAAGAACTGTAACACTCACTGCAAGGGTCCGCGGC 145

Qy        982 TTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGT 1041
Db        144 TTCATTCTTGAAATCAGCCAGACCAAGAACCCACTGGAAGGAACCAATTCCAGACACATT 85

Qy       1042 AGGAA 1046
Db         84 GCGAA 80
```

RESULT 44
AQ083508

LOCUS	AQ083508 600 bp DNA linear GSS 20-APR-1999
DEFINITION	RPCI11-52J19.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-52J19, genomic survey sequence.
ACCESSION	AQ083508
VERSION	AQ083508.1 GI:3444692
KEYWORDS	GSS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 600)
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE	Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL	Unpublished (1998)
COMMENT	Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html

Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .600
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="GDB:7519818"
	/db_xref="taxon:9606"
	/clone="RPCI-11-52J19"
	/sex="Male"
	/cell_type="Lymphocytes"
	/clone_lib="RPCI-11"
	/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"

ORIGIN

Query Match 16.7%; Score 333.8; DB 9; Length 600;
Best Local Similarity 77.7%; Pred. No. 5.7e-44;
Matches 439; Conservative 0; Mismatches 77; Indels 49; Gaps 1;

```
Qy        380 GTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGG 439
Db         32 GTCACCTAGGCTGGAGTGCTGTTGAGAGGTGAAGCTGTCTGGGCTCCTGGGTCGGGTGGG 91

Qy        440 GACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTG 499
Db         92 GCCTTGGAGAACTTTTCTGTCTAGCTAAAGGTTTGTAAATGCACCAATCAGCGCTCTGTG 151

Qy        500 TCTAGCTAAAGGATTGT---------------------------------------- 516
Db        152 TCTAGCTAAATGTTTGTAAACGCACCAATCAGCACTCTGTAAAAACAGACTGATCAGCAC 211

Qy        517 ------AAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGG 570
Db        212 TCTGTAAAATGGACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGAAGTGAGTGGG 271

Qy        571 GTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCC 630
Db        272 GCCAAATAAGGGAATAAAAGCTGGCCACCCAAGCCAGCAGCGGCAACCCGCTCAGGTCCC 331

Qy        631 CTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCAT 690
Db        332 CTTCTATGCTGTGGAAGCTTTGTTCTTTCGCTCTTCACAATAAATCTTGCTGCTGCTCAC 391

Qy        691 TCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCA 750
Db        392 TCTTTGGGGCCACACTACCTTTATGAGCTGTAACACTCACTGCGAAGGTCTGTGCTTCAC 451

Qy        751 TTCCTGAAGTCAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGC 810
Db        452 TCCTGAAGTCAGCAAGACCATGAACCCACCGGGAGGAACAAGTAACTCCGGATGCACCAC 511

Qy        811 CTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGA 870
Db        512 CTTTATGAGCTGTAACACTCACTACGAAGGTCTGCAGCTTCACTCCTGAAGTCAGTGAGA 571

Qy        871 CCACAAACCCACCAGAAGGAAGAAA 895
Db        572 CCATGAACCCACCAGGAGGGACAAA 596
```

RESULT 45
AQ507503/c

LOCUS	AQ507503 556 bp DNA linear GSS 29-APR-1999
DEFINITION	RPCI-11-298B4.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-298B4, genomic survey sequence.
ACCESSION	AQ507503
VERSION	AQ507503.1 GI:4712250
KEYWORDS	GSS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-298B4.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7614075"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-298B4"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             16.7%;  Score 333.2;  DB 9;  Length 556;
  Best Local Similarity   87.4%;  Pred. No. 7.3e-44;
  Matches  376;  Conservative    0;  Mismatches   53;  Indels    1;  Gaps    1;

Qy        614 CAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAA 673
              ||||||||||||||||||||||||| |||||| |||||||||||||||||||||||||||
Db        556 CAACCCACTCGGGTCCCCTTCCACGCTGTGGGAGCTTTGTTCTTTTGCTCTTCACAATAA 497

Qy        674 ATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGC 733
              |||||||||||||||| |||||| ||||||||| ||||||| |||||||||||| | |
Db        496 ATCTTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATCAGCTGTAACACTCGCCGT 437

Qy        734 GAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAA 792
              ||||||| |||||||||||||| |||||| | ||| ||| ||||||| |||||| ||||
Db        436 GAGGGTCCGTGGCTTCATTCCTCAAGTCAGCGAGATCACAAACCCACCAGAAGGAGCAAA 377

Qy        793 GAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCA 852
              |||||| ||    | ||||||||||||||||||||||||||||||| ||||||||||||
Db        376 TAACTCCAGACACACCACCTTTAAGAGCTGTAACACTCACTGCGAAGGTCTGCAGCTTCA 317

Qy        853 CTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAA 912
              |||||||||||| |||||||| |||| ||  |||||||||||||| ||||||| | |||
Db        316 CTCCTGAAGTCAGCGAGACCACGAACCTACTGGAAGGAAGAAACTCCGGACACATCCGAA 257

Qy        913 TATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGG 972
              ||||||||||||||| | |||||| ||||||||| |||||||||||||||||   | |
Db        256 CATCTGAAGGAACAAATTACAGACAAACCATCTTTAAGAGCTGTAACACTCACCATGAAG 197

Qy        973 GTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCC 1032
              |||   |||||||| | |||||||||||||||||||||||||| ||||||  |||||||
Db        196 GTCCACGGCTTCATCCCTGAAGTCAGCAAGACCAAGAACCCACCAGAAGGACTAAATTCC 137

Qy       1033 AGACACAGTA 1042
              |||||| ||
Db        136 GGACACAATA 127
```

```
Search completed: January 22, 2006, 21:09:21
Job time : 5433.54 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 01:00:32 ; Search time 6808.48 Seconds
(without alignments)
16706.187 Million cell updates/sec

Title: US-09-728-552A-3_COPY_59000_61000
Perfect score: 2001
Sequence: 1 aaaggcttctgaaatcagac.....cttgaagggttagagggact 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	39678	6	AX033938	AX033938 Sequence
	2	2001	100.0	80622	8	AF222855	AF222855 Homo sapi
	3	1999.4	99.9	39238	6	AX033939	AX033939 Sequence
	4	1999.4	99.9	69058	8	AF222856	AF222856 Homo sapi
	5	1999.4	99.9	80155	8	AF042484	AF042484 Homo sapi
	6	1999.4	99.9	80202	8	AF222854	AF222854 Homo sapi
	7	1999.4	99.9	196934	8	AL355530	AL355530 Human DNA
c	8	828	41.4	46324	8	AC093818	AC093818 Homo sapi
	9	810.6	40.5	131458	8	AC002386	AC002386 Homo sapi
c	10	799.8	40.0	195768	8	AC022034	AC022034 Homo sapi
c	11	794.4	39.7	62466	8	AC005613	AC005613 Homo sapi
c	12	793.6	39.7	107278	8	AC114962	AC114962 Homo sapi
	13	793.6	39.7	184536	14	AC068682	AC068682 Homo sapi
	14	785.2	39.2	25158	8	HSU116E9	Z95333 Human DNA s
	15	782	39.1	193041	8	AC092017	AC092017 Homo sapi
c	16	779	38.9	162397	14	AC011098	AC011098 Homo sapi
c	17	779	38.9	216031	8	AC011328	AC011328 Homo sapi
c	18	778	38.9	169482	8	AL390836	AL390836 Human DNA
	19	773.4	38.7	115231	8	HS785G19	AL035458 Human DNA
c	20	773.4	38.7	182753	14	AL137253	AL137253 Homo sapi
	21	772.6	38.6	174084	14	AC084395	AC084395 Homo sapi
	22	771.8	38.6	126595	8	AL160399	AL160399 Human DNA
	23	771.8	38.6	164532	14	AL355854	AL355854 Homo sapi
	24	771	38.5	172274	8	AP000924	AP000924 Homo sapi
	25	771	38.5	174311	8	AP002788	AP002788 Homo sapi
c	26	771	38.5	190592	14	AC015806	AC015806 Homo sapi
	27	761.2	38.0	168298	14	AC024401	AC024401 Homo sapi
	28	761.2	38.0	190018	8	AC013406	AC013406 Homo sapi
	29	754.4	37.7	145993	8	AC098972	AC098972 Homo sapi
	30	754.4	37.7	158508	8	AC107622	AC107622 Homo sapi
c	31	751.8	37.6	94768	14	BX571893	BX571893 Homo sapi
	32	751.8	37.6	99242	8	BX927235	BX927235 Human DNA
	33	751.8	37.6	104114	8	CR788250	CR788250 Human DNA
	34	751.8	37.6	109682	8	BX120007	BX120007 Human DNA
c	35	751.8	37.6	197067	14	BX571806	BX571806 Homo sapi
c	36	745.6	37.3	150705	8	AC084277	AC084277 Homo sapi
	37	745.6	37.3	184449	14	AC145142	AC145142 Homo sapi
c	38	742.2	37.1	75852	8	AC074133	AC074133 Homo sapi
c	39	742.2	37.1	183494	8	AC091977	AC091977 Homo sapi
	40	741.2	37.0	159797	8	AC104697	AC104697 Homo sapi
c	41	740.6	37.0	117583	8	AC034238	AC034238 Homo sapi
c	42	740.6	37.0	133322	14	AC091845	AC091845 Homo sapi
	43	740.4	37.0	136031	8	AL645568	AL645568 Human DNA
	44	739.8	37.0	203076	8	AL591074	AL591074 Human DNA
c	45	733.4	36.7	166357	8	AC026307	AC026307 Homo sapi
	46	733.4	36.7	185281	14	AC011214	AC011214 Homo sapi
c	47	733.4	36.7	313264	14	AC023053	AC023053 Homo sapi
c	48	731.8	36.6	150746	8	AC126567	AC126567 Homo sapi
	49	731.8	36.6	209258	14	AC019234	AC019234 Homo sapi
	50	726.4	36.3	108000	8	AP002878	AP002878 Homo sapi
	51	726.4	36.3	124686	14	AC087111	AC087111 Homo sapi
c	52	726.4	36.3	167035	14	AC027116	AC027116 Homo sapi
c	53	726.4	36.3	185388	14	AC009547	AC009547 Homo sapi
c	54	726.4	36.3	186562	14	AC011768	AC011768 Homo sapi
c	55	726.2	36.3	102218	8	AC124016	AC124016 Homo sapi
	56	726.2	36.3	139514	14	AC012233	AC012233 Homo sapi
c	57	726.2	36.3	188030	14	AC015456	AC015456 Homo sapi
	58	724.4	36.2	165764	14	AC022527	AC022527 Homo sapi
c	59	724.4	36.2	166358	8	AC091919	AC091919 Homo sapi
c	60	723	36.1	172387	14	AC073579	AC073579 Homo sapi
c	61	721.8	36.1	78835	8	AL390754	AL390754 Human DNA
c	62	716	35.8	154653	8	AC146050	AC146050 Pan trogl
	63	713.6	35.7	124047	8	AL713966	AL713966 Human DNA
	64	711.8	35.6	168012	14	AC025565	AC025565 Homo sapi
	65	710.6	35.5	126887	8	AL353146	AL353146 Human DNA
c	66	710	35.5	203726	8	AC084871	AC084871 Homo sapi
	67	709.4	35.5	130819	8	AL662842	AL662842 Human DNA
c	68	709.4	35.5	131234	8	HS172K2	Z84814 Human DNA s
	69	709.4	35.5	349980	6	CS039423	CS039423 Sequence
	70	708.4	35.4	134060	8	AC010389	AC010389 Homo sapi
	71	706.6	35.3	154315	8	AP005380	AP005380 Homo sapi
	72	706.6	35.3	157568	8	AP005138	AP005138 Homo sapi
c	73	706.6	35.3	158121	14	AC026594	AC026594 Homo sapi
	74	706.6	35.3	178676	14	AC026965	AC026965 Homo sapi
	75	706.6	35.3	185684	14	AC027481	AC027481 Homo sapi
c	76	705.4	35.3	172945	8	AC108099	AC108099 Homo sapi
c	77	705.4	35.3	307380	14	AC087155	AC087155 Mus muscu
	78	704.8	35.2	195768	8	AC022034	AC022034 Homo sapi
	79	703.8	35.2	259865	14	AC022972	AC022972 Homo sapi
	80	702.6	35.1	118771	8	AL929581	AL929581 Human DNA
	81	700.8	35.0	127820	8	HSJ612B15	AL049597 Human DNA
c	82	700.2	35.0	165004	14	AC060826	AC060826 Homo sapi
	83	699	34.9	127707	8	AC115222	AC115222 Homo sapi
c	84	699	34.9	194103	14	AC060818	AC060818 Homo sapi
	85	698.6	34.9	167487	8	AC148697	AC148697 Macaca mu
	86	696.8	34.8	169247	14	AC079124	AC079124 Homo sapi
c	87	696.8	34.8	190436	8	AC074182	AC074182 Homo sapi
	88	694.4	34.7	176843	14	AC026655	AC026655 Homo sapi
	89	692.2	34.6	163673	8	AC093240	AC093240 Homo sapi
	90	689.4	34.5	171741	14	AC025032	AC025032 Homo sapi
	91	687.8	34.4	151514	8	HS471M13	Z97198 Human DNA s

	92	687.2	34.3	164486	8	CNS05TES	AL359332 Human chr
c	93	684.4	34.2	106938	8	AC132032	AC132032 Homo sapi
c	94	684.2	34.2	165090	8	HS44A20	AL035086 Human DNA
	95	684	34.2	144665	8	CNS07EF7	AL512358 Human chr
c	96	683.8	34.2	71032	14	AC084084	AC084084 Homo sapi
c	97	683.4	34.2	103188	8	AL731551	AL731551 Human DNA
c	98	682.4	34.1	94219	8	AY796203	AY796203 Homo sapi
c	99	682.4	34.1	115177	8	AL583836	AL583836 Human DNA
c	100	682.2	34.1	130724	8	AC004253	AC004253 Homo sapi

ALIGNMENTS

RESULT 1
AX033938
LOCUS AX033938 39678 bp DNA linear PAT 21-SEP-2000
DEFINITION Sequence 30 from Patent WO9851790.
ACCESSION AX033938
VERSION AX033938.1 GI:10280507
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified
unclassified sequences.
REFERENCE 1
AUTHORS Cancilla,M.R., Choo,K.H. and Du,S.D.
TITLE A novel nucleic acid molecule
JOURNAL Patent: WO 9851790-A 30 19-NOV-1998;
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES Location/Qualifiers
source 1. .39678
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN

Query Match 100.0%; Score 2001; DB 6; Length 39678;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18083 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 18142

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18143 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 18202

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18203 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 18262

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18263 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 18322

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18323 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 18382

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18383 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 18442

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18443 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 18502

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18503 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 18562

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18563 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 18622

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18623 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 18682

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18683 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 18742

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18743 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 18802

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18803 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 18862

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18863 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 18922

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18923 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 18982

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18983 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 19042

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19043 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 19102

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19103 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 19162

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19163 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 19222

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19223 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 19282

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19283 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 19342

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19343 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 19402

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19403 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 19462

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19463 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 19522

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19523 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 19582

Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19583 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 19642

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19643 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 19702

Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19703 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 19762

Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19763 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 19822

Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19823 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 19882

Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19883 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 19942

Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19943 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 20002

Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20003 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 20062

Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001
                |||||||||||||||||||||
Db      20063 CCTTGAAGGGTTAGAGGGACT 20083
```

```
RESULT 2
AF222855
LOCUS       AF222855               80622 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222855
VERSION     AF222855.1  GI:9246845
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80622)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 80622)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Genomic sequence from human 10q25.2, clonelib=HC
FEATURES             Location/Qualifiers
     source          1. .80622
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="HC"
     repeat_region   complement(3. .176)
                     /rpt_family="Alu"
     repeat_region   complement(979. .3564)
                     /rpt_family="Tigger1"
     repeat_region   2967. .3094
                     /rpt_family="Alu"
     repeat_region   3664. .3934
                     /rpt_family="Alu"
     repeat_region   complement(4920. .5181)
                     /rpt_family="Alu"
     repeat_region   complement(6287. .6579)
                     /rpt_family="Alu"
     repeat_region   6625. .6966
                     /rpt_family="MLT2B2"
     repeat_region   complement(9215. .9490)
                     /rpt_family="Alu"
     repeat_region   complement(11280. .11565)
                     /rpt_family="Alu"
     repeat_region   complement(12428. .12675)
                     /rpt_family="Alu"
     repeat_region   complement(14304. .14583)
                     /rpt_family="Alu"
     repeat_region   17503. .17794
                     /rpt_family="Alu"
     repeat_region   complement(20250. .20533)
                     /rpt_family="Alu"
     repeat_region   complement(23031. .23323)
                     /rpt_family="Alu"
     repeat_region   complement(24301. .24479)
                     /rpt_family="MIR"
     repeat_region   24951. .25250
                     /rpt_family="Alu"
     repeat_region   complement(25987. .26342)
                     /rpt_family="THE1"
     repeat_region   27097. .27335
                     /rpt_family="MER2"
     repeat_region   27538. .27814
                     /rpt_family="Alu"
     repeat_region   27891. .28161
                     /rpt_family="Alu"
     repeat_region   29063. .29334
                     /rpt_family="Alu"
     repeat_region   complement(31025. .31093)
                     /rpt_family="MLT1"
     repeat_region   complement(31817. .32352)
                     /rpt_family="MER1"
     repeat_region   complement(32489. .32607)
                     /rpt_family="MLT1"
     repeat_region   33039. .33316
                     /rpt_family="Alu"
     repeat_region   33469. .33760
                     /rpt_family="Alu"
     repeat_region   complement(34527. .34590)
                     /rpt_family="MIR"
     repeat_region   34689. .34995
                     /rpt_family="Alu"
     repeat_region   complement(36349. .36437)
                     /rpt_family="MIR"
     repeat_region   complement(38796. .39079)
                     /rpt_family="Alu"
     repeat_region   39441. .39693
                     /rpt_family="Alu"
     repeat_region   complement(43176. .43237)
                     /rpt_family="MIR"
     repeat_region   complement(45203. .45478)
                     /rpt_family="Alu"
     repeat_region   45892. .46025
                     /rpt_family="MIR"
     repeat_region   complement(46618. .46851)
                     /rpt_family="Alu"
     repeat_region   48623. .48939
                     /rpt_family="Alu"
     repeat_region   complement(49048. .50698)
                     /rpt_family="L1"
     repeat_region   complement(50710. .50972)
                     /rpt_family="Alu"
     repeat_region   complement(51031. .51192)
                     /rpt_family="L1"
     repeat_region   complement(51450. .51739)
```

```
                     /rpt_family="Alu"
     repeat_region   complement(53270. .53562)
                     /rpt_family="Alu"
     repeat_region   54261. .54808
                     /rpt_family="LTR12"
     repeat_region   59460. .60054
                     /rpt_family="LTR12"
     repeat_region   60349. .60633
                     /rpt_family="Alu"
     repeat_region   complement(60669. .60801)
                     /rpt_family="MIR"
     repeat_region   complement(61684. .61852)
                     /rpt_family="MER5"
     repeat_region   63466. .63753
                     /rpt_family="Alu"
     repeat_region   64063. .64228
                     /rpt_family="MER5"
     repeat_region   64446. .64521
                     /rpt_family="MIR"
     repeat_region   65794. .65883
                     /rpt_family="MER5"
     repeat_region   66772. .66951
                     /rpt_family="MER20"
     repeat_region   complement(67895. .68416)
                     /rpt_family="MER41"
     repeat_region   complement(73192. .73470)
                     /rpt_family="Alu"
     repeat_region   74086. .74367
                     /rpt_family="Alu"
     repeat_region   75916. .76183
                     /rpt_family="Alu"
     repeat_region   complement(77771. .78059)
                     /rpt_family="Alu"
ORIGIN

  Query Match             100.0%;  Score 2001;  DB 8;  Length 80622;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59027 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 59086

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59087 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 59146

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59147 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 59206

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59207 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 59266

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59267 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 59326

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59327 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 59386

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59387 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 59446

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59447 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 59506

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59507 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 59566

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59567 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 59626

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59627 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 59686

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59687 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 59746

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59747 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 59806

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59807 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 59866

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59867 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 59926

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59927 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 59986

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59987 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 60046

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60047 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 60106

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60107 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 60166

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60167 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 60226

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60227 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 60286

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60287 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 60346

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60347 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 60406

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60407 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 60466

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60467 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 60526

Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60527 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 60586

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60587 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 60646
```

```
Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60647 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 60706

Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60707 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 60766

Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60767 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 60826

Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60827 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 60886

Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60887 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 60946

Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60947 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 61006

Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001
             |||||||||||||||||||||
Db      61007 CCTTGAAGGGTTAGAGGGACT 61027
```

```
RESULT 3
AX033939
LOCUS       AX033939               39238 bp    DNA     linear   PAT 21-SEP-2000
DEFINITION  Sequence 31 from Patent WO9851790.
ACCESSION   AX033939
VERSION     AX033939.1  GI:10280508
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Cancilla,M.R., Choo,K.H. and Du,S.D.
  TITLE     A novel nucleic acid molecule
  JOURNAL   Patent: WO 9851790-A 31 19-NOV-1998;
            CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
            DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES             Location/Qualifiers
     source          1. .39238
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             99.9%;  Score 1999.4;  DB 6;  Length 39238;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2000;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17759 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 17818

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17819 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 17878

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17879 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 17938

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17939 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 17998
```

```
Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17999 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 18058

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18059 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 18118

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18119 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 18178

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18179 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 18238

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18239 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 18298

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18299 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 18358

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18359 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 18418

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18419 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 18478

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18479 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 18538

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18539 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 18598

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18599 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 18658

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18659 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 18718

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18719 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 18778

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18779 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 18838

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18839 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 18898

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18899 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 18958

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18959 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 19018

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19019 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 19078

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
```

```
Db      19079 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACAAGGTC 19138

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440

Db      19139 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 19198

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500

Db      19199 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 19258

Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560

Db      19259 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 19318

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620

Db      19319 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 19378

Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680

Db      19379 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 19438

Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740

Db      19439 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 19498

Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800

Db      19499 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 19558

Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860

Db      19559 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 19618

Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920

Db      19619 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 19678

Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980

Db      19679 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 19738

Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001

Db      19739 CCTTGAAGGGTTAGAGGGACT 19759
```

```
RESULT 4
AF222856
LOCUS       AF222856               69058 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone PnC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222856
VERSION     AF222856.1  GI:9246846
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 69058)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 69058)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Human genomic sequence from 10q25.2, clonelib=PnC
FEATURES             Location/Qualifiers
     source          1. .69058
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="PnC"
     repeat_region   complement(598. .3183)
                     /rpt_family="Tigger1"
     repeat_region   2587. .2715
                     /rpt_family="Alu"
     repeat_region   3283. .3553
                     /rpt_family="Alu"
     repeat_region   complement(4548. .4809)
                     /rpt_family="Alu"
     repeat_region   complement(5919. .6211)
                     /rpt_family="Alu"
     repeat_region   6257. .6599
                     /rpt_family="MLT2B2"
     repeat_region   complement(8848. .9123)
                     /rpt_family="Alu"
     repeat_region   complement(10913. .11198)
                     /rpt_family="Alu"
     repeat_region   complement(12062. .12313)
                     /rpt_family="Alu"
     repeat_region   complement(13957. .14236)
                     /rpt_family="Alu"
     repeat_region   17133. .17424
                     /rpt_family="Alu"
     repeat_region   complement(19882. .20197)
                     /rpt_family="Alu"
     repeat_region   complement(22664. .22956)
                     /rpt_family="Alu"
     repeat_region   complement(23935. .24113)
                     /rpt_family="MIR"
     repeat_region   24585. .24884
                     /rpt_family="Alu"
     repeat_region   complement(25618. .25973)
                     /rpt_family="THE1"
     repeat_region   26732. .26970
                     /rpt_family="MER2"
     repeat_region   27173. .27449
                     /rpt_family="Alu"
     repeat_region   27526. .27796
                     /rpt_family="Alu"
     repeat_region   28702. .28978
                     /rpt_family="Alu"
     repeat_region   complement(30668. .30736)
                     /rpt_family="MLT1"
     repeat_region   complement(31460. .31993)
                     /rpt_family="MER1"
     repeat_region   complement(32130. .32248)
                     /rpt_family="MLT1"
     repeat_region   32685. .32961
                     /rpt_family="Alu"
     repeat_region   33113. .33403
                     /rpt_family="Alu"
     repeat_region   complement(34169. .34233)
                     /rpt_family="MIR"
     repeat_region   34332. .34639
                     /rpt_family="Alu"
     repeat_region   complement(35993. .36081)
                     /rpt_family="MIR"
     repeat_region   complement(38438. .38721)
                     /rpt_family="Alu"
     repeat_region   39081. .39333
                     /rpt_family="Alu"
     repeat_region   complement(42817. .42877)
                     /rpt_family="MIR"
     repeat_region   complement(44845. .45119)
```

```
                     /rpt_family="Alu"
     repeat_region   45532. .45665
                     /rpt_family="MIR"
     repeat_region   complement(46257. .46490)
                     /rpt_family="Alu"
     repeat_region   48262. .48578
                     /rpt_family="Alu"
     repeat_region   complement(48687. .50337)
                     /rpt_family="L1"
     repeat_region   complement(50349. .50611)
                     /rpt_family="Alu"
     repeat_region   complement(50670. .50831)
                     /rpt_family="L1"
     repeat_region   complement(51085. .51374)
                     /rpt_family="Alu"
     repeat_region   complement(52905. .53197)
                     /rpt_family="Alu"
     repeat_region   53896. .54467
                     /rpt_family="LTR12"
     repeat_region   59129. .59723
                     /rpt_family="LTR12"
     repeat_region   60018. .60302
                     /rpt_family="Alu"
     repeat_region   complement(60338. .60470)
                     /rpt_family="MIR"
     repeat_region   complement(61353. .61521)
                     /rpt_family="MER5"
     repeat_region   63135. .63416
                     /rpt_family="Alu"
     repeat_region   63724. .63889
                     /rpt_family="MER5"
     repeat_region   64107. .64182
                     /rpt_family="MIR"
     repeat_region   65459. .65548
                     /rpt_family="MER5"
     repeat_region   66439. .66618
                     /rpt_family="MER20"
     repeat_region   complement(67562. .68084)
                     /rpt_family="MER41"
ORIGIN

  Query Match             99.9%;  Score 1999.4;  DB 8;  Length 69058;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2000;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58696 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 58755

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58756 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 58815

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58816 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 58875

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58876 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 58935

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58936 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 58995

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58996 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 59055

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59056 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 59115

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59116 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 59175

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59176 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 59235

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59236 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 59295

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59296 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 59355

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59356 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 59415

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59416 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 59475

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59476 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 59535

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59536 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 59595

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59596 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 59655

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59656 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 59715

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59716 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 59775

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59776 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 59835

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59836 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 59895

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59896 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 59955

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59956 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 60015

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      60016 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACAAGGTC 60075

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60076 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 60135

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60136 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 60195

Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60196 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 60255

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60256 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 60315

Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60316 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 60375

Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60376 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 60435

Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60436 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 60495

Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60496 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 60555

Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60556 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 60615

Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60616 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 60675

Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001
                |||||||||||||||||||||
Db      60676 CCTTGAAGGGTTAGAGGGACT 60696
```

```
RESULT 5
AF042484
LOCUS       AF042484               80155 bp    DNA     linear   PRI 22-JUN-2000
DEFINITION  Homo sapiens sequence of the 80 kb core centromere protein binding
            region of the mar del(10) neocentromere corresponding to human
            chromosome 10q25.2, complete sequence.
ACCESSION   AF042484
VERSION     AF042484.1  GI:4205782
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,K.H.
  TITLE     Sequence analysis of an 80 kb human neocentromere
  JOURNAL   Hum. Mol. Genet. 8 (2), 217-227 (1999)
   PUBMED   9931329
REFERENCE   2  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville, Melbourne, Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .80155
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /note="Sequence information obtained from BAC clones
                     prepared from cell line BE2C1-18-5f"
     repeat_region   complement(join(49. .525,3628. .4545,4849. .5053))
                     /rpt_family="L1MEc"
                     /rpt_type=dispersed
     repeat_region   join(560. .597,3184. .3281)
                     /rpt_family="MER58"
                     /rpt_type=dispersed
     repeat_region   complement(join(598. .2553,2716. .3183))
                     /rpt_family="TIGGER1"
                     /rpt_type=dispersed
     repeat_region   2558. .2581
                     /standard_name="T1"
                     /note="100% homology"
                     /rpt_type=direct
                     /rpt_unit="(caggcacagtgg)2"
     repeat_region   2570. .2715
                     /rpt_family="AluSg/x"
                     /rpt_type=dispersed
     repeat_region   3021. .3050
                     /standard_name="T2"
                     /note="83.3% homology"
                     /rpt_type=direct
                     /rpt_unit="(taacaaagtg)3"
     repeat_region   3283. .3558
                     /rpt_family="AluSx"
                     /rpt_type=dispersed
     repeat_region   4173. .4211
                     /standard_name="T3"
                     /note="82.1% homology"
                     /rpt_type=direct
                     /rpt_unit="(aaaaaaataattt)3"
     repeat_region   complement(4546. .4848)
                     /rpt_family="AluSx"
                     /rpt_type=dispersed
     repeat_region   complement(5909. .6208)
                     /rpt_family="AluSg"
                     /rpt_type=dispersed
     repeat_region   join(6232. .6573,6638. .6688)
                     /rpt_family="MLT2CB"
                     /rpt_type=dispersed
     repeat_region   6598. .6645
                     /standard_name="AC24"
                     /note="100% homology"
                     /rpt_type=tandem
                     /rpt_unit="ac"
     repeat_region   6598. .6637
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ca"
     repeat_region   7091. .7134
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   7808. .7911
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   8349. .8524
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   complement(8839. .9124)
                     /rpt_family="AluJb"
                     /rpt_type=dispersed
     repeat_region   9180. .9280
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   10070. .10170
                     /note="low complexity"
                     /rpt_family="purine-rich"
     repeat_region   10097. .10136
                     /standard_name="GA1"
                     /note="82.5% homology"
                     /rpt_type=direct
                     /rpt_unit="(gaaagaaagg)4"
     repeat_region   10175. .10339
                     /rpt_family="FRAM"
                     /rpt_type=dispersed
     repeat_region   complement(10899. .11197)
                     /rpt_family="AluSx"
```

```
                /rpt_type=dispersed
misc_feature    join(11627. .11757,12434. .12512,13443. .13562)
                /note=">98% identity"
                /db_xref="dbEST:AA774571"
repeat_region   11996. .12028
                /standard_name="AT4"
                /note="100% homology"
                /rpt_type=tandem
                /rpt_unit="(attt)8"
repeat_region   complement(12010. .12312)
                /rpt_family="AluJb"
                /rpt_type=dispersed
repeat_region   12327. .12433
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(13946. .14235)
                /rpt_family="AluSq"
                /rpt_type=dispersed
repeat_region   15130. .15159
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   15164. .15753
                /standard_name="AT28"
                /note="82.1% homology"
                /rpt_type=direct
                /rpt_unit="(atgtatatatgtgtatatagacataaat)21"
repeat_region   join(15170. .15289,15410. .15612)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="ta"
repeat_region   15617. .15731
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="cata"
repeat_region   16004. .16100
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16128. .16364)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16407. .16592)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16657. .16889)
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   16914. .16936
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   17102. .17386
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   17406. .17564
                /standard_name="GA2"
                /note="81.8% homology"
                /rpt_type=direct
repeat_region   17416. .17543
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   17647. .17702
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   complement(19863. .20161)
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   complement(20842. .20913)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(21366. .21414)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   21375. .21639
                /standard_name="CT1"
                /note="75% homology"
                /rpt_type=dispersed
repeat_region   complement(21415. .21536)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="tggaa"
repeat_region   complement(21524. .21646)
                /note="low complexity"
                /rpt_family="purine-rich"
repeat_region   22409. .22646
                /rpt_family="L1M4"
                /rpt_type=dispersed
repeat_region   22560. .22604
                /standard_name="T4"
                /note="80% homology"
                /rpt_type=direct
                /rpt_unit="(aatattacaataatt)3"
```

Query Match 99.9%; Score 1999.4; DB 8; Length 80155;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58676 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 58735

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58736 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 58795

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58796 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 58855

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58856 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 58915

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58916 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 58975

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58976 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 59035

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59036 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 59095

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59096 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 59155

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59156 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 59215

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59216 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 59275

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59276 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 59335

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59336 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 59395

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59396 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 59455

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59456 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 59515

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59516 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 59575

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59576 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 59635

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59636 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 59695

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59696 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 59755

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59756 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 59815

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59816 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 59875

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59876 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 59935

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59936 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 59995

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59996 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACAAGGTC 60055

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60056 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 60115

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60116 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 60175

Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60176 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 60235

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60236 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 60295

Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60296 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 60355

Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60356 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 60415

Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60416 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 60475

Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      60476 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 60535

Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60536 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 60595

Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60596 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 60655

Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001
                |||||||||||||||||||||
Db      60656 CCTTGAAGGGTTAGAGGGACT 60676
```

```
RESULT 6
AF222854
LOCUS       AF222854               80202 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222854
VERSION     AF222854.1  GI:9246844
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80202)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 80202)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Human genomic sequence from 10q25.2, clonelib=NC, second release
FEATURES             Location/Qualifiers
     source          1. .80202
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="NC"
     repeat_region   complement(598. .3183)
                     /rpt_family="Tigger1"
     repeat_region   2587. .2715
                     /rpt_family="Alu"
     repeat_region   3283. .3553
                     /rpt_family="Alu"
     repeat_region   complement(4548. .4809)
                     /rpt_family="Alu"
     repeat_region   complement(5919. .6211)
                     /rpt_family="Alu"
     repeat_region   6257. .6599
                     /rpt_family="MLT2B2"
     repeat_region   complement(8848. .9123)
                     /rpt_family="Alu"
     repeat_region   complement(10913. .11198)
                     /rpt_family="Alu"
     repeat_region   complement(12062. .12313)
                     /rpt_family="Alu"
     repeat_region   complement(13957. .14236)
                     /rpt_family="Alu"
     repeat_region   17133. .17424
                     /rpt_family="Alu"
     repeat_region   complement(19882. .20197)
```

```
                     /rpt_family="Alu"
     repeat_region   complement(22664. .22956)
                     /rpt_family="Alu"
     repeat_region   complement(23935. .24113)
                     /rpt_family="MIR"
     repeat_region   24585. .24884
                     /rpt_family="Alu"
     repeat_region   complement(25618. .25973)
                     /rpt_family="THE1"
     repeat_region   26732. .26970
                     /rpt_family="MER2"
     repeat_region   27173. .27449
                     /rpt_family="Alu"
     repeat_region   27526. .27796
                     /rpt_family="Alu"
     repeat_region   28702. .28978
                     /rpt_family="Alu"
     repeat_region   complement(30668. .30736)
                     /rpt_family="MLT1"
     repeat_region   complement(31460. .31993)
                     /rpt_family="MER1"
     repeat_region   complement(32130. .32248)
                     /rpt_family="MLT1"
     repeat_region   32685. .32961
                     /rpt_family="Alu"
     repeat_region   33113. .33403
                     /rpt_family="Alu"
     repeat_region   complement(34169. .34233)
                     /rpt_family="MIR"
     repeat_region   34332. .34639
                     /rpt_family="Alu"
     repeat_region   complement(35993. .36081)
                     /rpt_family="MIR"
     repeat_region   complement(38438. .38721)
                     /rpt_family="Alu"
     repeat_region   39081. .39333
                     /rpt_family="Alu"
     repeat_region   complement(42817. .42877)
                     /rpt_family="MIR"
     repeat_region   complement(44845. .45119)
                     /rpt_family="Alu"
     repeat_region   45532. .45665
                     /rpt_family="MIR"
     repeat_region   complement(46257. .46490)
                     /rpt_family="Alu"
     repeat_region   48262. .48578
                     /rpt_family="Alu"
     repeat_region   complement(48687. .50337)
                     /rpt_family="L1"
     repeat_region   complement(50349. .50611)
                     /rpt_family="Alu"
     repeat_region   complement(50670. .50831)
                     /rpt_family="L1"
     repeat_region   complement(51085. .51374)
                     /rpt_family="Alu"
     repeat_region   complement(52905. .53197)
                     /rpt_family="Alu"
     repeat_region   53896. .54467
                     /rpt_family="LTR12"
     repeat_region   59129. .59719
                     /rpt_family="LTR12"
     repeat_region   60018. .60302
                     /rpt_family="Alu"
     repeat_region   complement(60338. .60470)
                     /rpt_family="MIR"
     repeat_region   complement(61353. .61521)
                     /rpt_family="MER5"
     repeat_region   63135. .63416
                     /rpt_family="Alu"
     repeat_region   63724. .63889
                     /rpt_family="MER5"
     repeat_region   64107. .64182
                     /rpt_family="MIR"
     repeat_region   65459. .65548
                     /rpt_family="MER5"
     repeat_region   66439. .66618
                     /rpt_family="MER20"
     repeat_region   complement(67562. .68084)
                     /rpt_family="MER41"
     repeat_region   complement(72861. .73139)
                     /rpt_family="Alu"
     repeat_region   73756. .74037
                     /rpt_family="Alu"
     repeat_region   75588. .75853
                     /rpt_family="Alu"
     repeat_region   complement(77441. .77729)
                     /rpt_family="Alu"
ORIGIN

  Query Match             99.9%;  Score 1999.4;  DB 8;  Length 80202;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2000;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58696 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 58755

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58756 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 58815

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58816 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 58875

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58876 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 58935

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58936 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 58995

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58996 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 59055

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59056 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 59115

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59116 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 59175

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59176 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 59235

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59236 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 59295

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59296 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 59355

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59356 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 59415

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59416 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 59475

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59476 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 59535

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59536 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 59595

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59596 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 59655

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59656 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 59715

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59716 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 59775

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59776 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 59835

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59836 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 59895

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59896 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 59955

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59956 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 60015

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      60016 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACAAGGTC 60075

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60076 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 60135

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60136 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 60195

Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60196 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 60255

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60256 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 60315

Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60316 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 60375

Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60376 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 60435

Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60436 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 60495

Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60496 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 60555

Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      60556 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 60615

Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60616 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 60675

Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001
                |||||||||||||||||||||
Db      60676 CCTTGAAGGGTTAGAGGGACT 60696
```

```
RESULT 7
AL355530
LOCUS       AL355530              196934 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-338L11 on chromosome 10 Contains
            the 5' end of the gene for novel gene (KIAA0534) and a CpG island,
            complete sequence.
ACCESSION   AL355530
VERSION     AL355530.6  GI:8894326
KEYWORDS    HTG; KIAA0534.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 196934)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 1, 2000 this sequence version replaced gi:8670617.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            RP11-338L11 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .196934
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-338L11"
                     /clone_lib="RPCI-11.2"
     misc_feature    1
                     /note="Clone_left_end: RP11-338L11"
     gene            join(68829. .69507,95654. .95737,97168. .97281,
                     103062. .103190,104794. .105002,135506. .135680,
                     141023. .141110,146500. .>146837)
                     /gene="RP11-537G20.1"
                     /locus_tag="RP11-537G20.1-002"
```

```
gene            join(68829. .69100,72107. .72301,95654. .95737,97168. .97281,
                103062. .103190,103876. .103977,104794. .105002,
                135506. .135680,141023. .141110,146500. .146837)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-003"
mRNA            join(68829. .69507,95654. .95737,97168. .97281,
                103062. .103190,104794. .105002,135506. .135680,
                141023. .141110,146500. .>146837)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-002"
                /product="novel protein"
                /note="match: cDNAs: BC035157.1"
mRNA            join(68829. .69100,72107. .72301,95654. .95737,97168. .97281,
                103062. .103190,103876. .103977,104794. .105002,
                135506. .135680,141023. .141110,146500. .146837)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-002"
                /product="novel protein"
                /note="match: cDNAs: BC029592.1"
CDS             join(69215. .69507,95654. .95737,97168. .97281,
                103062. .103190,104794. .105002,135506. .135680,
                141023. .141110,146500. .>146837)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-002"
                /standard_name="OTTHUMP00000020556"
                /codon_start=1
                /product="novel protein"
                /protein_id="CAI13482.1"
                /db_xref="GI:55958328"
                /db_xref="InterPro:IPR000859"
                /db_xref="InterPro:IPR006209"
                /db_xref="InterPro:IPR006210"
                /db_xref="InterPro:IPR006652"
                /db_xref="InterPro:IPR011498"
                /db_xref="UniProt/TrEMBL:Q5T5Y9"
                /translation="METGGRARTGTPQPAAPGVWRARPAGGGGGGASSWLLDGNSWLL
                CYGFLYLALYAQVSQSKPCERTGSCFSGRCVNSTCLCDPGWVGDQCQHCQGRFKLTEP
                SGYLTDGPINYKYKTKCTWLIEGYPNAVLRLRFNHFATECSWDHMYVYDGDSIYAPLI
                AVLSGLIVPEIRGNETVPEVVTTSGYALLHFFSDAAYNLTGFNIFYSINSCPNNCSGH
                GKCTTSVSVPSQVYCECDKYWKGEACDIPYCKANCGSPDHGYCDLTGEKLCVCNDSWQ
                GPDCSLNVPSTESYWILPNVKPFSPSVGRASHKAVLHGKFMWVIGGYTFNYSSFQMVL
                NYNLESSIWNVGTPSRGPLQRYGHSLALYQENIFMYGGRIETNDGNVTDELWVFNIHS
                QSWSTKTPTVLGHGQQYAVEGHSAHIMELDSRDVVMIIIFGYSAIYGYTSSIQEYHIC
                ELLKNCNFFIDWECFSLIKSSYEFSQ"
gene            join(<141072. .141110,146500. .146755,191160. .191343,
                AL512304.10:22131. .22285,AL512304.10:45441. .45525,
                AL512304.10:47045. .47253,AL512304.10:48574. .48692,
                AL512304.10:61636. .61768,AL512304.10:66497. .66678,
                AL357059.13:11678. .11891,AL357059.13:13499. .13687,
                AL357059.13:27162. .27380,AL357059.13:45926. .46063,
                AL356100.8:34176. .34269,AL356100.8:65767. .65819,
                AL356100.8:101458. .101552,AL356100.8:106691. .106769,
                AL356100.8:108689. .108846,AL356100.8:158780. .158841,
                AL356100.8:188975. .189053,AL392087.7:178400. .178507,
                AC091667.9:114936. .115050,AC022542.5:69593. .73927)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-001"
mRNA            join(<141072. .141110,146500. .146755,191160. .191343,
                AL512304.10:22131. .22285,AL512304.10:45441. .45525,
                AL512304.10:47045. .47253,AL512304.10:48574. .48692,
                AL512304.10:61636. .61768,AL512304.10:66497. .66678,
                AL357059.13:11678. .11891,AL357059.13:13499. .13687,
                AL357059.13:27162. .27380,AL357059.13:45926. .46063,
                AL356100.8:34176. .34269,AL356100.8:65767. .65819,
                AL356100.8:101458. .101552,AL356100.8:106691. .106769,
                AL356100.8:108689. .108846,AL356100.8:158780. .158841,
                AL356100.8:188975. .189053,AL392087.7:178400. .178507,
                AC091667.9:114936. .115050,AC022542.5:69593. .73927)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-001"
                /product="novel protein"
                /note="match: ESTs: AI588945.1 AI817331.1 R37016.1
                match: cDNAs: AB011106.2 BC027764.1"
CDS             join(<141072. .141110,146500. .146755,191160. .191343,
                AL512304.10:22131. .22285,AL512304.10:45441. .45525,
                AL512304.10:47045. .47253,AL512304.10:48574. .48692,
                AL512304.10:61636. .61768,AL512304.10:66497. .66678,
                AL357059.13:11678. .11891,AL357059.13:13499. .13687,
                AL357059.13:27162. .27380,AL357059.13:45926. .46063,
                AL356100.8:34176. .34269,AL356100.8:65767. .65819,
                AL356100.8:101458. .101552,AL356100.8:106691. .106769,
                AL356100.8:108689. .108846,AL356100.8:158780. .158841,
                AL356100.8:188975. .189053,AL392087.7:178400. .178507,
                AC091667.9:114936. .115050,AC022542.5:69593. .69714)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-001"
                /standard_name="OTTHUMP00000020555"
                /codon_start=1
                /product="novel protein"
                /protein_id="CAI13483.1"
                /db_xref="GI:55958329"
                /db_xref="GOA:Q5VV63"
                /db_xref="InterPro:IPR001304"
                /db_xref="InterPro:IPR002049"
                /db_xref="InterPro:IPR002165"
                /db_xref="InterPro:IPR003659"
                /db_xref="InterPro:IPR006652"
                /db_xref="InterPro:IPR011498"
                /db_xref="UniProt/TrEMBL:Q5VV63"
                /translation="PLQRYGHSLALYQENIFMYGGRIETNDGNVTDELWVFNIHSQSW
                STKTPTVLGHGQQYAVEGHSAHIMELDSRDVVMIIIFGYSAIYGYTSSIQEYHISSNT
                WLVPETKGAIVQGGYGHTSVYDEITKSIYVHGGYKALPGNKYGLVDDLYKYEVNTKTW
                TILKESGFARYLHSAVLINGAMLIFGGNTHNDTSLSNGAKCFSADFLAYDIACDEWKI
                LPKPNLHRDVNRFGHSAVVINGSMYIFGGFSSVLLNDILVYKPPNCKAFRDEELCKNA
                GPGIKCVWNKNHCESWESGNTNNILRAKCPPKTAASDDRCYRYADCASCTANTNGCQW
                CDDKKCISANSNCSMSVKNYTKCHVRNEQICNKLTSCKSCSLNLNCQWDQRQQECQAL
                PAHLCGEGWSHIGDACLRVNSSRENYDNAKLYCYNLSGNLASLTTSKEVEFVLDEIQK
                YTQQKVSPWVGLRKINISYWGWEDMSPFTNTTLQWLPGEPNDSGFCAYLERAAVAGLK
                ANPCTSMANGLVCEKPVVSPNQNARPCKKPCSLRTSCSNCTSNGMECMWCSSTKRCVD
                SNAYIISFPYGQCLEWQTATCSPQNCSGLRTCGQCLEQPGCGWCNDPSNTGRGHCIEG
                SSRGPMKLIGMHHSEMVLDTNLCPKEKNYEWSFIQCPACQCNGHSTCINNNVCEQCKN
                LTTGKQCQDCMPGYYGDPTNGGQCTACTCSGHANICHLHTGKCFCTTKGIKGDQCQLC
                DSENRYVGNPLRGTCYYSLLIDYQFTFSLLQEDDRHHTAINFIANPEQSNKNLDISIN
                ASNNFNLNITWSVGSTAGTISGEETSIVSKNNIKEYRDSFSYEKFNFRSNPNITFYVY
                VSNFSWPIKIQIAFSQHNTIMDLVQFFVTFFSCFLSLLLVAAVVWKIKQTCWASRRRE
                QLLRERQQMASRPFASVDVALEVGAEQTEFLRGPLEGAPKPIAIEPCAGNRAAVLTVF
                LCLPRGSSGAPPPGQSGLAIASALIDISQQKASDSKDKTSGVRNRKHLSTRQGTCV"
misc_feature    196934
                /note="Clone_right_end: RP11-338L11"
ORIGIN
```

```
  Query Match             99.9%;  Score 1999.4;  DB 8;  Length 196934;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2000;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17198 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 17257

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17258 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 17317

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17318 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 17377

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17378 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 17437

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17438 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 17497

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
```

```
Db      17498 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 17557
Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
Db      17558 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 17617
Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
Db      17618 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 17677
Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
Db      17678 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 17737
Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
Db      17738 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 17797
Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
Db      17798 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 17857
Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
Db      17858 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 17917
Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
Db      17918 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 17977
Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
Db      17978 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 18037
Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
Db      18038 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 18097
Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
Db      18098 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 18157
Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
Db      18158 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 18217
Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
Db      18218 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTTCAGGGTTAC 18277
Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
Db      18278 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 18337
Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
Db      18338 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 18397
Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
Db      18398 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 18457
Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
Db      18458 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 18517
Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
Db      18518 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 18577
Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
Db      18578 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 18637
Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500
Db      18638 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 18697
Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
Db      18698 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 18757
Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
Db      18758 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 18817
Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680
Db      18818 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 18877
Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740
Db      18878 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 18937
Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800
Db      18938 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 18997
Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860
Db      18998 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 19057
Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920
Db      19058 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 19117
Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980
Db      19118 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 19177
Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001
Db      19178 CCTTGAAGGGTTAGAGGGACT 19198
```

```
RESULT 8
AC093818/c
LOCUS       AC093818               46324 bp    DNA     linear   PRI 21-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-366L15 from 2, complete sequence.
ACCESSION   AC093818 AC020681
VERSION     AC093818.3  GI:18497233
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 46324)
  AUTHORS   Kozlowicz,A., Doebber,A. and Nguyen,C.
  TITLE     The sequence of Homo sapiens BAC clone RP11-366L15
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 46324)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 46324)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 46324)
  AUTHORS   Waterston,R.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 46324)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 46324)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 5, 2002 this sequence version replaced gi:15809168.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0366L15
                Drafting Center: WIBR
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP11-343H10.  Actual start of
            this clone is at base position 1 of RP11-366L15.

            The clone RP11-366L15 contains a 150 bp insertion not sequenced in
            the right neighboring clone, RP11-343H10. This insertion sequence
            is being submitted as H_NH0366L15_F1.

            The sequence of AC020681 has been incorporated into AC093818.
FEATURES              Location/Qualifiers
     source           1. .46324
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="2"
                      /clone="RP11-366L15"
                      /clone_lib="RPCI-11"
     misc_feature     4019. .5261
                      /note="CpG island (%GC=69.8, o/e=0.89, #CpGs=143)"
     gene             4654. .41594
                      /gene="PDK1"
     mRNA             join(4654. .4892,7254. .7395,10766. .10837,13049. .13233,
                      13524. .13619,15402. .15479,17287. .17363,19272. .19370,
                      34824. .34934,41481. .41594)
                      /gene="PDK1"
     CDS              join(4697. .4892,7254. .7395,10766. .10837,13049. .13233,
                      13524. .13619,15402. .15479,17287. .17363,19272. .19370,
                      34824. .34934,41481. .>41594)
                      /gene="PDK1"
                      /note="Homo sapiens pyruvate dehydrogenase kinase,
                      isoenzyme 1, mRNA (cDNA clone MGC:24867 IMAGE:4778360),
                      complete cds.; H_NH0366L15.1
                      This gene was based on gi(24660127)
                      Continues as H_NH0343H10.1"
                      /codon_start=1
                      /product="unknown"
                      /protein_id="AAY14915.1"
                      /db_xref="GI:62822366"
                      /translation="MRLARLLRGAALAGPGPGLRAAGFSRSFSSDSGSSPASERGVPG
                      QVDFYARFSPSPLSMKQFLDFGSVNACEKTSFMFLRQELPVRLANIMKEISLLLPDNLL
                      RTPSVQLVQSWYIQSLQELLDFKDKSAEDAKAIYDFTDTVIRIRNRHNDVIPTMAQGV
                      IEYKESFGVDPVTSQNVQYFLDRFYMSRISIRMLLNQHSLLFGGKGKGSPSHRKHIGS
                      INPNCNVLEVIKDGYENARRLCDLYYINSPELELEELNAKSPGQPIQVVYVPSHLYHM
                      VFELFKNAMRATMEHHANRGVYPPIQVHVTLGNEDLTVKMSDRGGGVPLRKIDRLFNY
                      MYSTAPRPRVETSRAVPLAGFGYGLPISRLYAQYFQGDLKLYSLEGYGTDAVIYIK"
ORIGIN
```

```
  Query Match             41.4%;  Score 828;  DB 8;  Length 46324;
  Best Local Similarity   88.3%;  Pred. No. 2.4e-209;
  Matches  922;  Conservative    0;  Mismatches  120;  Indels    2;  Gaps    2;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||| |||| |||||||||||||  |||||||||||||||||||||||||||||||||||
Db      39867 AAAGTCTTCCGAAATCAGACAACACCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 39808

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||||||||||||||||||||||| || || | || ||
Db      39807 AACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTATTACTCACCTTCGG 39748

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||||||||| ||| ||||||||||||||
Db      39747 ACCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGAATCAAGGCCATCAAGCT 39688

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              | ||||| |||||||||||| || ||||||||| |||||||||||||||||    ||||||
Db      39687 ATAGATGGTCTTACAAATGGAATCCCAAATGAACTCAACTAACAACTTCTATCAAGGACC 39628

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||| |||||||||||||||||| |||||||||||| ||||||||||||||||||||
Db      39627 CCTGGACTGACCCGCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 39568

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              |||||||||||||||||| |||||||||||| |||||||||||||||||||||||| ||
Db      39567 AACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAAGAAGTAGCTACAGCGGTCATCGCCC 39508

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||| ||| |||||| | |||||| |||||||||| |||||||||||||| |
Db      39507 AATTCCCAACAACAGTTGGGGTATACTGTTTAGAGGGGGGATGGAGAGGTGAAGCCAGTT 39448

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||| |||||||||||||||||||| |||||||||||||||||||||||||
Db      39447 GGGCTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAAAGGATTGTAAAT 39388

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT 539
              |||||||||||||||||||||||||||||||||||| |||||||  ||||||||||||||||||
Db      39387 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGTTTGTAAACACACCAATCAGCACTCTGT 39328

Qy        540 AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACC 599
              |||||||||||||||||||| |||| |||| |||||||||| ||||| |||||||||
Db      39327 AAAATGGACCAATCAGCAGGACATGGGTGGGGCCAAATAAGGGAATAAAAGCTGGCCACC 39268

Qy        600 CGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTT 659
              | |||||||||||||||| |||||||||||||||||||| |||| ||||||||||
Db      39267 CCAGCCAGCAGTGGCAACCCGCTCGGGTCCCCTTCCACACTGCGGAAACTTTGTTCTTTC 39208
```

```
Qy        660 GCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCT 719
              ||||||||||||||||||| |||||||||| |||||| ||| |||||| ||||||||||
Db      39207 ACTCTTCACAATAAATCTTACTGCTGCTCACTCTTTGGGTCTGCACTATCTTTATGAGCC 39148

Qy        720 GTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCA 778
              |||||||||||  ||| ||||||  |||||| |||||||||| | |||||| |||||||
Db      39147 GTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGTCAGCGAGACCATGAACCCA 39088

Qy        779 CTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAA 838
              |  | ||||||||| |||||| ||  |||  ||||||||||||||||||||||||||||
Db      39087 CCAGGAGGAACAAACAACTCCAGACATGCCACCTTTAAGAGCTGTAACACTCACTGCGAA 39028

Qy        839 GCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTC 898
              | ||||  |||||||||||||| ||||   |   |   || ||  | |||| |||||||||
Db      39027 GGTCTGTGGCTTCACTCCTGAGGTCAAGCAAGACCACAAACCCACCGAAGAAAGAAACTC 38968

Qy        899 TGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAA 958
              |||||||| || || ||||||||||||||||||||||||||||||| | | ||||||
Db      38967 TGGACACATCTAAACATCTGAAGGAACAAACTCCAGACACACCATCTTTAAAAACTGTAA 38908

Qy        959 CACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGG 1018
              ||||||||   |||||| |  |||||||||  |||||||||||||||||||| ||| |
Db      38907 CACTCACCATGAGGGTCCGCAGCTTCATTCTCAAAGTCAGCAAGACCAAGAACCTACCAG 38848

Qy       1019 AAGGAACAAATTCCAGACACAGTA 1042
              ||| ||| |||||| || |  |||
Db      38847 AAGAAACCAATTCCGGATATGGTA 38824
```

```
RESULT 9
AC002386
LOCUS       AC002386              131458 bp    DNA     linear   PRI 08-OCT-2003
DEFINITION  Homo sapiens BAC clone CTA-343H22 from 7, complete sequence.
ACCESSION   AC002386
VERSION     AC002386.2  GI:19034024
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 131458)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 131458)
  AUTHORS   Cordes,M.
  TITLE     The sequence of Homo sapiens BAC clone CTA-343H22
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 131458)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 131458)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 131458)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 131458)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 131458)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   8  (bases 1 to 131458)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 1, 2002 this sequence version replaced gi:2275181.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_RG343H22
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project and the Washington University Genome Sequencing Center. For
            additional information about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            Clone CTA-343H22 is from a release of the human BAC library
            CITB-HS-A. The library contains cloned DNA from human sperm. See:
            Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
            Kim et al., Genomics 34:213-8 (1996). The clone is available from
            Research Genetics, Inc. (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP11-573B16.  Actual start of
            this clone is at base position 1 of CTA-343H22 actual end is at
            base position 131458 of CTA-343H22.
FEATURES                Location/Qualifiers
     source             1. .131458
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="7"
                        /map="7"
                        /clone="CTA-343H22"
                        /clone_lib="CITB-HS-A"
     repeat_region      32. .236
                        /rpt_family="ERV1"
     repeat_region      256. .523
                        /rpt_family="MaLR"
     repeat_region      1137. .1227
                        /rpt_family="ERV1"
     repeat_region      1228. .1485
                        /rpt_family="Alu"
     repeat_region      1486. .1897
                        /rpt_family="ERV1"
     repeat_region      2406. .2489
                        /rpt_family="Tip100"
     repeat_region      2721. .2877
```

```
                /rpt_family="L1"
repeat_region   3370. .3614
                /rpt_family="L1"
repeat_region   3619. .4195
                /rpt_family="L1"
repeat_region   4790. .5239
                /rpt_family="ERVL"
repeat_region   5697. .5736
                /rpt_family="(GA)n"
repeat_region   6172. .6629
                /rpt_family="L1"
repeat_region   6630. .6662
                /rpt_family="(TA)n"
repeat_region   6664. .6698
                /rpt_family="(TG)n"
repeat_region   6699. .6805
                /rpt_family="L1"
repeat_region   6806. .7175
                /rpt_family="MaLR"
repeat_region   7178. .8788
                /rpt_family="MaLR"
repeat_region   8789. .9159
                /rpt_family="MaLR"
repeat_region   9161. .9573
                /rpt_family="L1"
repeat_region   9971. .10070
                /rpt_family="Alu"
repeat_region   10094. .10263
                /rpt_family="MaLR"
repeat_region   10420. .10461
                /rpt_family="AT_rich"
repeat_region   10764. .10881
                /rpt_family="L2"
repeat_region   10882. .11191
                /rpt_family="Alu"
repeat_region   11197. .11391
                /rpt_family="Alu"
repeat_region   11392. .11828
                /rpt_family="L2"
repeat_region   11851. .11954
                /rpt_family="ERVL"
repeat_region   12429. .13851
                /rpt_family="L1"
repeat_region   14181. .14202
                /rpt_family="AT_rich"
repeat_region   15058. .15223
                /rpt_family="MER1_type"
repeat_region   15238. .15523
                /rpt_family="Alu"
repeat_region   16058. .16320
                /rpt_family="Alu"
repeat_region   16328. .16758
                /rpt_family="ERVK"
repeat_region   16896. .17262
                /rpt_family="L1"
repeat_region   17405. .17867
                /rpt_family="L1"
repeat_region   20694. .21124
                /rpt_family="ERVL"
repeat_region   21138. .21261
                /rpt_family="AcHobo"
repeat_region   22074. .22304
                /rpt_family="MER2_type"
repeat_region   22381. .22732
                /rpt_family="L1"
repeat_region   23389. .23835
                /rpt_family="ERVL"
repeat_region   24144. .24174
                /rpt_family="(TA)n"
repeat_region   26614. .26822
                /rpt_family="MIR"
repeat_region   27691. .28064
                /rpt_family="MaLR"
repeat_region   28148. .28260
                /rpt_family="L1"
repeat_region   28280. .28373
                /rpt_family="CT-rich"
repeat_region   28416. .28528
                /rpt_family="L2"
repeat_region   29982. .30048
                /rpt_family="L2"
repeat_region   30453. .30778
                /rpt_family="ERVL"
repeat_region   30782. .30902
                /rpt_family="L1"

  Query Match             40.5%;  Score 810.6;  DB 8;  Length 131458;
  Best Local Similarity   87.7%;  Pred. No. 9.9e-205;
  Matches  938;  Conservative    0;  Mismatches  109;  Indels   22;  Gaps    4;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db      73714 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAGCTCTTATACCAACCTCTGGAGTTGGGC 73773

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||| |||||||||||| |||||||||||||| || | || || ||
Db      73774 AACATGGCTTCTCCCCTTGCTAGGTCCTGTGGCAGCCATCTTGCTATTACTTGCCTTCGG 73833

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||||||||||| ||||||||||||
Db      73834 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATCAAGGCCATCAAGCT 73893

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db      73894 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTACTGAGGACA 73953

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||| ||||| |||||||||||| |||||||||||| |||| ||||||||||||||
Db      73954 CCTGGACTGACCCACTGGCCCTTTCACTGGCCTAAAGAGTTCCCTTCTGGAGGACACTAC 74013

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              |||||||||||||| ||||||||||||||||||||||||||||| | | ||||| | ||
Db      74014 AACTGCAGGGCCCCGTCTTCACCCCTATCCAGCAGGAAGTAGCTAGATCAGTCATTGCCC 74073

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||||||| |||||||||||||||||||||||||||||
Db      74074 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCAGCT 74133

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              | ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db      74134 GAGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 74193

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT 539
              |||||||||| |||||||                |  ||||||||||||||||||||||
Db      74194 GCACCAATCATCACTCTG----------------TAAAAATGCACCAATCAGCACTCTGT 74237

Qy        540 AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGG----C 595
              |||||||||||||||| |||| |||||| ||| |||||||  | ||||| |||     |
Db      74238 AAAATGGACCAATCAGTAGGACGTGGGCAGGGCCAAATAAGACAATAAAAGCTGTCCACC 74297

Qy        596 CACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTC 655
              ||||||||||||||  ||||||  ||| |||||||||||  ||||||||||||||||||
Db      74298 CACCCGAGCCAGCAGCAGCAACCTGCTCAGGTCCCCTTCCATGCTGTGGAAGCTTTGTTC 74357

Qy        656 TTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATG 715
              |||||||||||||||||||||||||||| ||||| ||| |||||| |||  |||||||||||
Db      74358 TTTTGCTCTTCACAATAAATCTTGATGCTGATCACTCTTTGGGTCTGCACTACCTTTATG 74417

Qy        716 AGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCACGAA 774
              ||||||||||||||  |||||||||||||||||||||||||||||   ||||||||||
Db      74418 AGCTGTAACACTCACCACGAGGGTCTGTGGCTTCATTCCTGAAGTCAGTGAGACCACGAA 74477

Qy        775 CCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTG 834
              ||||  | |||||||| |||||| ||    |  ||||||||||||||||||||||||||
Db      74478 TCCACCAGGAGGAACAAACAACTCCAGACACACCACCTTTAAGAGCTGTAACACTCACTG 74537
```

```
Qy        835 CGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAA 894
              |||| ||||| |||||||||||||||||||| |||||||| |||||||| ||||||||||
Db      74538 TGAAGGTCTGCGGCTTCACTCCTGAAGTCAGCGAGACCACGAACCCACCGGAAGGAAGAA 74597

Qy        895 ACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCT 954
              ||||  |||||| ||||| |||||||||||||||  |  |||||||||||||| ||||||
Db      74598 ACTCCAGACACATCTGAACATCTGAAGGAACAAACCCTGGACACACCATCTTTAAGAGCT 74657

Qy        955 GTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCA 1014
              |||||||||||| |||||||||||| |||||||||||||||||||  ||||||||||||||
Db      74658 GTAACACTCACTGCAAGGGTCTGTAGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCA 74717

Qy       1015 CCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCT 1063
              || ||||||| ||||||  |||||| ||  | ||   |  ||||  ||
Db      74718 CCAGAAGGAATAAATTCTGGACACATTATCACATGGATTATTTTAGCCT 74766
```

RESULT 10
AC022034/c

```
LOCUS       AC022034              195768 bp    DNA     linear   PRI 01-JAN-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-400K9, complete sequence.
ACCESSION   AC022034
VERSION     AC022034.7  GI:17426331
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 195768)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-400K9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195768)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 195768)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 8, 2001 this sequence version replaced gi:17386392.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2101
                Center clone name: 400_K_9
            --------------.
FEATURES             Location/Qualifiers
     source          1. .195768
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-400K9"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   914. .1047
                     /rpt_family="MIR3"
     repeat_region   1634. .1928
                     /rpt_family="AluSg"
     repeat_region   complement(5008. .5192)
                     /rpt_family="MER5A"
     repeat_region   complement(5686. .5778)
                     /rpt_family="MIR"
     repeat_region   complement(5817. .5972)
                     /rpt_family="MIR3"
     repeat_region   6807. .6894
                     /rpt_family="(TG)n"
     repeat_region   7044. .7299
                     /rpt_family="MIR"
     repeat_region   complement(8103. .8247)
                     /rpt_family="FLAM_C"
     repeat_region   complement(9261. .9553)
                     /rpt_family="AluSq"
     repeat_region   10329. .10358
                     /rpt_family="AT_rich"
     repeat_region   10794. .10901
                     /rpt_family="MER96B"
     repeat_region   11864. .11884
                     /rpt_family="AT_rich"
     repeat_region   12585. .12658
                     /rpt_family="MER113"
     repeat_region   complement(12696. .12900)
                     /rpt_family="L1MC5"
     repeat_region   12901. .13207
                     /rpt_family="AluSx"
     repeat_region   complement(13208. .13252)
                     /rpt_family="L1MC5"
     repeat_region   13321. .13718
                     /rpt_family="MER66A"
     repeat_region   complement(13719. .14021)
                     /rpt_family="AluSx"
     repeat_region   14022. .14091
                     /rpt_family="MER66A"
     repeat_region   complement(14192. .14356)
                     /rpt_family="L1MC5"
     repeat_region   complement(14418. .15549)
                     /rpt_family="L1MC5"
     repeat_region   complement(15545. .15682)
                     /rpt_family="L1MD1"
```

```
repeat_region   complement(15685. .17136)
                /rpt_family="L1MD2"
repeat_region   17137. .17440
                /rpt_family="AluJb"
repeat_region   complement(17441. .18005)
                /rpt_family="L1MD2"
repeat_region   complement(18002. .19359)
                /rpt_family="L1MD2"
repeat_region   19384. .19509
                /rpt_family="MIR"
repeat_region   complement(19606. .19983)
                /rpt_family="MER113"
repeat_region   19990. .20204
                /rpt_family="L1MC4a"
repeat_region   20205. .20238
                /rpt_family="(CA)n"
repeat_region   20396. .20601
                /rpt_family="L1MDa"
repeat_region   complement(20602. .20753)
                /rpt_family="THE1C"
repeat_region   complement(20756. .21978)
                /rpt_family="MSTA-int"
repeat_region   21979. .22016
                /rpt_family="(TTTA)n"
repeat_region   complement(22017. .22293)
                /rpt_family="AluSx"
repeat_region   complement(22294. .22709)
                /rpt_family="MSTA-int"
repeat_region   complement(22710. .23093)
                /rpt_family="MSTA"
repeat_region   23150. .24986
                /rpt_family="L1PA13"
repeat_region   24987. .25158
                /rpt_family="AluJb"
repeat_region   25429. .25719
                /rpt_family="AluSx"
repeat_region   26007. .26139
                /rpt_family="L2"
repeat_region   26406. .26531
                /rpt_family="L2"
repeat_region   complement(26607. .26716)
                /rpt_family="MER57A"
repeat_region   complement(27622. .27755)
                /rpt_family="MIR"
repeat_region   27775. .27877
                /rpt_family="L2"
repeat_region   complement(27929. .28003)
                /rpt_family="MIR"
repeat_region   complement(28066. .28275)
                /rpt_family="MSTA"
repeat_region   complement(28276. .31266)
                /rpt_family="L1PA12"
repeat_region   complement(31267. .31566)
                /rpt_family="AluSg"
repeat_region   complement(31567. .32933)
                /rpt_family="L1PA12"
repeat_region   complement(33237. .34054)
                /rpt_family="L1PA12"
repeat_region   complement(34153. .34631)
                /rpt_family="L1PA12"
repeat_region   complement(34637. .34770)
                /rpt_family="FLAM_C"
repeat_region   complement(35125. .35415)
                /rpt_family="AluSg"
repeat_region   35633. .35690
                /rpt_family="MIR"
repeat_region   35713. .36004
                /rpt_family="L2"
repeat_region   complement(36568. .36845)
                /rpt_family="L2"
repeat_region   complement(36852. .37702)
                /rpt_family="L1ME1"
repeat_region   complement(37711. .38494)
```

```
  Query Match             40.0%;  Score 799.8;  DB 8;  Length 195768;
  Best Local Similarity   73.0%;  Pred. No. 7.4e-202;
  Matches 1273;  Conservative    0;  Mismatches  267;  Indels  203;  Gaps   11;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||
Db     124698 AAAGGCTTCTGAAATGGGACAATGCCTTTCAAACTCTTATACCAATCTCTGGAGTTGGGC 124639

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||||  ||| ||||||||||||||| || | || |||||
Db     124638 AACATGGCTTCTCCCCTTTCTAGGTCACGTGGCAGCCATCTTGCTATTACTTGCCTTTGG 124579

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||||||||| |||||||||||||| ||
Db     124578 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGAATCGAGGCCATCAA-CT 124520

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||||||||||||| | ||||||
Db     124519 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTACCAAGGACC 124460

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||| |||| |||||||||||  || |||| |||||||||||||||||| |||||
Db     124459 CCTGGACTGACCTGCTGGCCCTTTCCCTGACCTAGAGAGCTCCCCTCTGGAGGATACTAC 124400

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||| | ||
Db     124399 AACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCAGTCATTGCCC 124340

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||||||||| ||| |||||||||||||||||||| |||
Db     124339 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGAGGGGATTGAGAGGTGAAGCCGGCT 124280

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||||| || |||||||||||||||||| |||||||||||| ||||||||
Db     124279 GGGCTTCTGGGTCGGGAGGGGACTTGGAGAACTTTTCTGTCTAGCTAAAGTATTGTAAAC 124220

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCA------ 533
              |||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db     124219 ACACCAATCAGCACTCTGTGTCTAGCTAAATGTTTGTAAATGCACCAATCAGCACTCTGT 124160

Qy        534 ------------------------------------------------------------ 533

Db     124159 AAAAACGCACCAATCAGCACTCTGTGTTTAGCTAAAGGTTTGTAAACGCACCAATCAGCA 124100

Qy        534 ----------------------------------------------CTCTGTAAAAT 544
                                                           ||||||||||
Db     124099 CTCTGTAAAAATGAACCAATCAGCATTTTGTAAAATGGACCAATCAGTGCTCTGTAAAAT 124040

Qy        545 GGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGC 604
              |||||| ||||||||    |||| |  | |||||||||| ||||| ||| ||||| ||||
Db     124039 GGACCAGTCAGCAGGCCATGGGTGAAGCCAAATAAGGGAATAAAAGCTGACCACCTGAGC 123980

Qy        605 CAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCT 664
              ||||||  ||||||  || |||||||||||||| |||||||| ||||||||||||||||
Db     123979 CAGCAGCAGCAACCTGCTTGGGTCCCCTTCCACGCTGTGGAAACTTTGTTCTTTTGCTCT 123920

Qy        665 TCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAAC 724
              |||||||||||||||||||||||| | |||| ||||  ||||||||||||||||||||||
Db     123919 TCACAATAAATCTTGCTGCTGCTCACTTTTTG-GTCCGCACTACCTTTATGAGCTGTAAC 123861

Qy        725 ACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGA 783
              ||||||  | ||| |||||||||||  ||||||||||    |||||||||||||||  |
Db     123860 ACTCACCACAAGGTTCTGTGGCTTCACTCCTGAAGTCAGGGAGACCACGAACCCACCAGG 123801

Qy        784 AGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCT 843
              ||||||||| |||||| |||||||  |||||||||||||||| |||||||||| |||
Db     123800 AGGAACAAACAACTCCAGATGTGCCACCTTTAAGAGCTGTAACATTCACTGCGAAGGTCT 123741

Qy        844 GCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGAC 903
              |  |||||||||||||||||  ||||||| ||||||||||||||||||||||||||||||
Db     123740 GTGGCTTCACTCCTGAAGTCAGCAAGACCACGAACCCACCAGAAGGAAGAAACTCTGGAC 123681
```

```
Qy        904 ACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTC 963
              ||| ||||| |||||||||||||||||||||||||||||||||||| ||| |||||||||
Db     123680 ACATCTGAACATCTGAAGGAACAAACTCCAGACACACCATCTTTAAGAATGGTAACACTC 123621

Qy        964 ACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGA 1023
              || |  |||||||||||||||| |||||||||||  |||||||||||||||| ||||||
Db     123620 ACTGTGAGGGTCTGTGGCTTCATCCTTGAAGTCAGTGAGACCAAGAACCCACCAGAAGGA 123561

Qy       1024 ACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACTCC 1083
              |  |||| |||||||| ||| ||||
Db     123560 ATCAATTACAGACACACTAGCAAAT---------------------------------- 123536

Qy       1084 AGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACATG 1143
                                          | |   || |  ||||||   | | ||      ||
Db     123535 -------------------------TAATTAAACCCAAGAATGAAGTAGTGGGAACTTTG 123501

Qy       1144 TGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCT 1203
                |    || |    |||   |       ||| ||   | ||| ||||        |
Db     123500 ATTAATAGCTTGTCAGTCAGAAGTATAGATGACAGAGCCTACTTGTGATTGGTGTCTGAA 123441

Qy       1204 GGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTT 1263
                | ||              ||     |  ||| |  |   ||  | || | |
Db     123440 ATGGGGAG-------------AAGTCTGGTGGGACTGAGTCCTTAACCTGTGGGATCTGA 123394

Qy       1264 TGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCGAT 1323
              |||   |   || |   |        || ||| ||||       |||  ||| || ||  |
Db     123393 TGCCATTTCCACGTAGTGTCAGAACAGAAATAAATT-------ATAAGGCCGGGCGCAGT 123341

Qy       1324 GGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCAGG 1383
              |||||||||||||||||  |||| ||||||||  |||       ||||||||||||||||
Db     123340 AGCTCATGCCTGTAATCCCGACACTCTGGGAGGCTGAGGTTATCAGATCACGAGGTCAGG 123281

Qy       1384 CGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAA--TACAAAAA 1441
              | | ||||||| ||||||  |||| |||||||| | ||||||||||||||     |||||
Db     123280 AGATTGAGACCATCCTGGCTAACACGGTGAAACTCTGTCTCTACTAAAAAATACAAAAAA 123221

Qy       1442 TTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGGCAGGGG 1500
              |||||| || |||||||| | | ||||| |||||| ||||| ||||||||||||||| |
Db     123220 TTAGCCAGGCGTGGTGGCGGGTACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAG 123161

Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
              |||  | |||||   |||||| ||| |||||||||| ||||||||  |||| ||||||||
Db     123160 AATGGCCTGAACCCAGGAGGCGGAGCTTGCAGTGAG-CAAGATCATGCCACTGCACTCCA 123102

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
              |||||||||||||| ||||||||||||||||||| |||||||| |||  |||  ||||
Db     123101 GCCTGGGTGACAGAGTGAGACTCTGTCTCAAAAAATAAAAAAAATAAAATAAAATAAAAA 123042

Qy       1621 ATA 1623
                ||
Db     123041 TTA 123039
```

```
RESULT 11
AC005613/c
LOCUS       AC005613               62466 bp    DNA     linear   PRI 04-SEP-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 114k9 (LBNL H94), complete
            sequence.
ACCESSION   AC005613
VERSION     AC005613.1  GI:3540152
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 62466)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 62466)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 62466)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .62466
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q"
                     /clone="114k9"
                     /note="LBNL H94"
     misc_feature    3810. .3936
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   complement(4364. .4417)
                     /rpt_family="L1"
     repeat_region   6262. .6639
                     /rpt_family="MLT2B2"
     repeat_region   complement(7533. .7788)
                     /rpt_family="Alu"
     repeat_region   10336. .10468
                     /rpt_family="MIR"
     repeat_region   13065. .13197
                     /rpt_family="L1"
     repeat_region   15009. .15297
                     /rpt_family="Alu"
     misc_feature    complement(17115. .17237)
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   complement(17287. .17757)
                     /rpt_family="MER9"
     repeat_region   complement(18497. .19099)
                     /rpt_family="LTR12"
     misc_feature    complement(19289. .19376)
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   complement(22605. .22883)
                     /rpt_family="Alu"
     misc_feature    complement(24937. .25098)
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(26944. .27544)
                     /rpt_family="LTR12"
     repeat_region   complement(28172. .28391)
                     /rpt_family="MER43"
     misc_feature    28588. .28762
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   28866. .29046
                     /rpt_family="MER25"
     repeat_region   complement(30344. .30634)
                     /rpt_family="Alu"
     repeat_region   31121. .32400
                     /rpt_family="L1"
     repeat_region   32804. .33080
                     /rpt_family="Alu"
     repeat_region   34143. .34470
                     /rpt_family="L1"
     repeat_region   complement(37778. .41031)
                     /rpt_family="L1"
     misc_feature    complement(38485. .39555)
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   complement(41034. .41274)
                     /rpt_family="Alu"
     repeat_region   42715. .42951
                     /rpt_family="L1"
```

```
     repeat_region   47785. .47824
                     /note="(GTAT)10"
                     /rpt_type=tandem
                     /rpt_unit="gtat"
     repeat_region   complement(47832. .52964)
                     /rpt_family="L1"
     repeat_region   52127. .52323
                     /rpt_family="Alu"
     repeat_region   complement(52568. .52810)
                     /rpt_family="MER25"
     repeat_region   complement(52999. .53250)
                     /rpt_family="L1"
     repeat_region   complement(53678. .54006)
                     /rpt_family="Alu"
     misc_feature    complement(54423. .54468)
                     /note="GRAIL 2 excellent exon, frame 1"
     misc_feature    56193. .56322
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   57743. .57782
                     /note="(AC)20"
                     /rpt_type=tandem
                     /rpt_unit="ac"
     repeat_region   59472. .59636
                     /rpt_family="MER5"
     repeat_region   complement(60588. .60746)
                     /rpt_family="L1"
     repeat_region   61139. .61455
                     /rpt_family="Alu"
ORIGIN

  Query Match             39.7%;  Score 794.4;  DB 8;  Length 62466;
  Best Local Similarity   86.0%;  Pred. No. 2.1e-200;
  Matches  937;  Conservative    0;  Mismatches  126;  Indels   27;  Gaps    4;

Qy          2 AAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGCG 61
              |||||||||||||||||||| |||||||||||||||||||||||||||||||| ||||||
Db      19537 AAGGCTTCTGAAATCAGACAAAGCCTTTCAAACTCTTATACCAACCTCTGGAATTGGGCA 19478

Qy         62 ACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGGG 121
              ||||||||||||||||||||||||||  || |||||||||||||| || || | ||||||
Db      19477 ACATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTATTACTCACCTTTGGG 19418

Qy        122 CCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCTA 181
              |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db      19417 CCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCTA 19358

Qy        182 CAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACCC 241
              |||||||||||||||| |||||||||||||| |||||| |||||| | |||||||||||
Db      19357 CAGATGATCTTACAAATGGAACCCCAAATGAGCTTAACTAATGACTTCTACCAAGGACCC 19298

Qy        242 CTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTACCC 301
              |||||| |||  |||||||||||||| ||||||| |||| |||||||||||| |||||||
Db      19297 CTGGACTGACTTGCTGGCCCTTTCACTGGCCTAGAGAGTTCCCCTCTGGAGAACACTACA 19238

Qy        302 ACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCCA 360
              |||| |||||||||||||| |||||||||||||||||||||||| |  |||||||  |||
Db      19237 ACTGTAGGGCCCCTTCTTTACCCCTATCCAGCAGGAAGTAGCTAGAATGGTCATCACCCA 19178

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              | |||||||||||||||||| |||||||| |||| ||||| |||||||||||||| ||||
Db      19177 ATTCCCAACAGCAGCTGGGGTCTCCTGTTTAGAGGAGGGATGGAGAGGTGAAGCCGGCTG 19118

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              || ||||||||  |||| |||||||| ||||||| | ||||||||||||||||||||
Db      19117 GGATTCTGGGTTGGGTGAGGACTTGGAAAACTTTTCTTTCTAGCTAAAGGATTGTAAACA 19058

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTG-- 538
              |||||||||| |||||||||||||||||||| ||||||| ||||||||||||||||||
Db      19057 CACCAATCAGCGCTCTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATCAGCACTCTGAA 18998

Qy        539 ----------------------TAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAA 576
                                   |||||||||||||||| |||   |||| |||| ||||
Db      18997 AAATGGACCAATCAGTGCTCTATAAAATGGACCAATCAGTAGGGCATGGGTGGGGCCAAA 18938

Qy        577 TAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA 636
              ||||||| ||||| ||||||||| ||||||||| ||||||  ||||||||||||||||||
Db      18937 TAAGGGAATAAAAGCTGGCCACCTGAGCCAGCAGCAGCAACCTGCTCGGGTCCCCTTCCA 18878

Qy        637 CACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTG 696
              || |||||||||||||||||| |||||||||||||||||||||| ||||||| ||||||
Db      18877 CAGTGTGGAAGCTTTGTTCTTTCGCTCTTCACAATAAATCTTGCTACTGCTCACTCTTTG 18818

Qy        697 TGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTG 756
              |||| |||||||||||||||||| |||||||||| | ||||||  |||||| |||||
Db      18817 GGTCCGCACTACCTTTATGAGCTGTGACACTCACTGCAAAGGTCTGCAGCTTCACTCCTG 18758

Qy        757 AAGTCAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAA 816
              ||||| | |||||| |||||||||| ||||||||| |||||   |    |  |||||||
Db      18757 AAGTC-AGAGACCATGAACCCACTGGGAGGAACAAACAACTCTGAACACACCACCTTTAA 18699

Qy        817 GAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTC-AGTGAGACCACA 875
              |||||||||||||||||||||||| ||||  |||||||| |||||||| ||||||||||
Db      18698 GAGCTGTAACACTCACTGCGAAGGTCTGTGGCTTCACTGCTGAAGTCAAGTGAGACCACG 18639

Qy        876 AACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGA 935
              |||||||||||||||||||||||| ||||||| ||||| |||||||||| ||||||||||
Db      18638 AACCCACCAGAAGGAAGAAACTCCGGACACATCTGAACATCTGAAGGAATAAACTCCAGA 18579

Qy        936 CACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGT 995
              |||||||||||| ||| |||||||||||||||| ||||||    |||||||||||||||
Db      18578 CACACCATCTTTAAGAACTGTAACACTCACCGCGAGGGTCCACTGCTTCATTCTTGAAGT 18519

Qy        996 CAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATT 1055
              |||| ||||||||||||||||  ||||||| |||||  |||||   |  | | |||
Db      18518 CAGCGAGACCAAGAACCCACCAAAAGGAACCAATTCTGGACACGCCACCACAGCTGAGAA 18459

Qy       1056 TTTGATCTGT 1065
              | || |  ||
Db      18458 TGTGCTGGGT 18449
```

```
RESULT 12
AC114962/c
LOCUS       AC114962              107278 bp    DNA     linear   PRI 05-NOV-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-206N2, complete sequence.
ACCESSION   AC114962
VERSION     AC114962.2  GI:24580362
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 107278)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 107278)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 107278)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 5, 2002 this sequence version replaced gi:19424423.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.7% of Sequence;
            Estimated Total Number of Errors is 0.2.
            NOTE:  This insert is not the entire sequence of the clone (entire
            sequence is 201.6kb). It is clipped at the overlaps with AC027342
```

```
          and AC010621. The number of bases overlapped with AC027342 is 5439
          and with AC010621 is 28528.
FEATURES             Location/Qualifiers
     source          1. .107278
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-206N2"
ORIGIN

  Query Match             39.7%;  Score 793.6;  DB 8;  Length 107278;
  Best Local Similarity   68.2%;  Pred. No. 3.4e-200;
  Matches 1339;  Conservative    0;  Mismatches  444;  Indels  179;  Gaps   10;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||  ||||||||  ||||||||||||  ||||||||||||
Db       6414 AAAGGCTTCTGAAATCAGACAACACCTTTCAAGCTCTTATACCAACTTCTGGAGTTGGGC 6355

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||  || ||||||||||||  || ||  | |||||
Db       6354 AACATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTGTTACTCACCTTTGG 6295

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||| ||||||||||||  || ||||||||||||||||| | ||| ||||||||||||
Db       6294 GCCCTATATTTTTAACCTTTTTGTCAAATTTGTTTCCTCTGGAATCAAGGCCATCAAGCC 6235

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              |||||||  |||||||| || ||| |||||||||  ||||  |||||||||| | ||||||||
Db       6234 ACAGATGGTCTTACAAGTGGAACACCAAATGAGTTCAACCAACAACTTCTACCGAGGACC 6175

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||  ||||  |||||||||  ||  ||||||||||||  ||||||||||||||  |||
Db       6174 CCTGGACTGACCTGCTGGCCCTTCCACTGGCCTAAAGAGTTCCCCTCTGGAGGACATTAC 6115

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG--C 358
              ||||||||||  |||||||  |||  ||||||||||||||||||||  ||||||||||  |
Db       6114 AACTGCAGGGCTCCTTCTTTGCCCTTATCCAGCAGGAAGTAGCTAAAGCGGTCATCGGCC 6055

Qy        359 CAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGC 418
              ||  ||||||||||||  |||||  ||||||||  ||||||||||||||||||||||||  ||
Db       6054 AAATTCCCAACAGCAGTTGGGGCATCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCGGC 5995

Qy        419 TGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              ||||||||||||   ||||||||||||||||| ||||  ||||||||||||||||||||||
Db       5994 TGGGCTTCTGGGCTGGGTGGGGACTTGGAGAA-TTTTCTGTCTAGCTAAAGGATTGTAAA 5936

Qy        479 TGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATT------------------------ 514
              ||||||||||||  |||||||||||||||| | ||
Db       5935 TGCACCAATCAGAGCTCTGTGTCTAGCTAAAAGGTTTGTAAACACACCAATCAGCACTCT 5876

Qy        515 -GTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTC 573
               ||||| ||||||||||  ||||||||||||||||||||||||||||||||||||||| |
Db       5875 GTAAAATGGACCAATCAGCTCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCC 5816

Qy        574 AAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTT 633
              |||||||||| |||||  |||||||||| || ||||||||||||||  ||| ||||||||||
Db       5815 AAATAAGGGAATAAAAGCTGGCCACCCAAGTCAGCAGTGGCAACCCGCTCAGGTCCCCTT 5756

Qy        634 CCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCT 693
              |||  ||||||||||||||||||||  ||||||  ||||||||||||||||||||||  |||
Db       5755 CCATGCTGTGGAAGCTTTGTTCTTTCACTCTTCGCAATAAATCTTGCTGCTGCTCACTCT 5696

Qy        694 TTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTC 753
              |||  ||||  |||||  |||  |||  |||||  || |  || ||||||  ||||||  ||
Db       5695 TTGGGTCCGCACTAACTTCATGAGTTGTCACACTTACCGTGAAGGTCTGCAGCTTCACTC 5636

Qy        754 CTGAAGTCA---------------------------------------------------- 762
              |||||| ||
Db       5635 CTGAAGCCACTGAGACCATGGACCCACCAGGAGGAAAAAGCAACTCCAGACGTGCCACCT 5576

Qy        763 ----------------------------------------------------ACAGACC 769
                                                                   |||||

Db       5575 TTAAGAGCTGTAACACTCACCGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACC 5516

Qy         770 ACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACT 829
               || |||||||||||||||||||| |||||| || |  |  ||||||||||||||||| ||
Db        5515 ACAAACCCACTGGAAGGAACAAACAACTCCTGACGCACCACCTTTAAGAGCTGTAACCCT 5456

Qy         830 CACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGG 889
               ||| | |||| |||| ||||||||||||||||||||| |||||||| ||||||||||||||
Db        5455 CACAGTGAAGGTCTGAAGCTTCACTCCTGAAGTCAGCGAGACCACGAACCCACCAGAAGG 5396

Qy         890 AAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCA 949
               ||||||||| ||| ||| ||||| |||||||||||||||||| |    ||||||||| |
Db        5395 AAGAAACTCCGGATACATCTGAACATCTGAAGGAACAAACTCCGG----ACCATCTTTAA 5340

Qy         950 GAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGA 1009
               || |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        5339 GAACTGTAACACTCACCGCAAGGGTCTGTGGCTTTGTTCTTGAAGTCAGCAAGACCAAGA 5280

Qy        1010 ACCCACCGGAAGGAACAAATTCCAGACACAGTAGG----------------------A 1045
               | |||| ||||||||  ||||||  ||||||  |||                    |
Db        5279 ATCCACTGGAAGGAATCAATTCCGGACACATAAGGCCGGGCACGGTGGCTCACGCCTGCA 5220

Qy        1046 AATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCA 1105
               | |      ||||   || |  |  | |  | | ||   ||  ||   |
Db        5219 ATCCCAGCACTTTGGGAGGTCGAGGCAGGTGGATCACGAGGTTAGGAGATCAAGACCATC 5160

Qy        1106 GCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGC 1165
                    | | ||  |  | ||   ||  |    || | ||  |
Db        5159 CTGGCTAACACAGTGAAACCCCGTCTCTACTGAAAATACAAAAACATTAGCCGAGCATGG 5100

Qy        1166 TTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATA 1225
               | |  |  |  |  | |  ||  ||  ||  || ||  ||  | ||||  |
Db        5099 TGGCAGTCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGCGTAAACCTG 5040

Qy        1226 ATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTT------GCATCTGAGACATGT 1279
                  |||        |  | || | |     || |     || |  | ||||
Db        5039 GGAGGCGGAGCTTGCAGTGAGCTGAGATTGTGCCACTGTACTCCAGCCTGGGTGACAGAG 4980

Qy        1280 AGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCGATGGCTCATGCCTGTAAT 1339
                ||  || | |    ||  | ||||||  |  ||  || | |||||||||  |||||||||
Db        4979 TGAGACTCCATCTCAAAAAAAAAAAAAAAAAAATGCTGGGCGTGATGGCTCACACCTGTAAT 4920

Qy        1340 CCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCAGGCGTTCGAGACCAGCCT 1399
               |||||||||||||||||| ||| | |  || ||| |||||||| | || |||||| |||
Db        4919 CCCAGCACTTTGGGAGGCCGAGGCGGGCAGACCACCAGGTCAGGAGATCAAGACCATCCT 4860

Qy        1400 GGCCAACATGGTGAAACCCCGTCTCTACTAAAAA------TACAAAAATTAGCCTGGTGT 1453
               |||||||||||||||||| |||| ||||||||      | ||||||||||| ||  |
Db        4859 GGCCAACATGGTGAAACCCCCTCTCCACTAAAAAAAATACAAAAAAAATTAGCCAGGCAT 4800

Qy        1454 GGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAAC 1512
               || ||   || |||||  |||  | ||||| |||||||  |||||||  ||||  | |||||
Db        4799 GGGGGTGAGCGTCTGTAGTCCTGGCTACTCCGGAGGCTAAGGCAGGAGAATGGCGTGAAC 4740

Qy        1513 TTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCATCCTGGGTGACA 1572
                ||||||  ||| ||||| ||||||  |||||||||||| ||||||||  ||||||  |||
Db        4739 CCGGGAGGTGGAGCTTGCAATGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCAACA 4680

Qy        1573 GAGCGAGACTCTGTCTCAAA------AAAAAAAAAAAAAAAAGGAAAGGAAAATATAATC 1626
               |||||||||| ||||||||       ||| ||| |||||| |  ||   | ||   |||
Db        4679 GAGCGAGACTCCGTCTCAAAGAAAATAAATAAATAAAAAATAAAAATAAATAAATAAATA 4620

Qy        1627 AAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACACTAAGT 1686
               || || ||  |  ||||  || |  | | || ||  |       |  |    ||
Db        4619 AAAAAGATAAAGTGGTAGGCTTAAGCCCAAACATATCAATAGTTACATTAAATGTAAATG 4560

Qy        1687 GTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGCTGTTAT 1746
               |||| ||    |  |  | |  |                || ||     |||  | ||
Db        4559 GTTTAACTATAAAAATTAAAATACAGTAACTAGCATAATGGATTAAAAACATGCCCTAAT 4500

Qy        1747 TATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGT 1788
               ||| |   || || |   ||   | | || ||||  || ||
Db        4499 TATATGTTCTCTACAAGAAACTCACATCAAATACAATAACGT 4458
```

```
RESULT 13
AC068682
LOCUS       AC068682              184536 bp    DNA     linear   HTG 23-SEP-2000
DEFINITION  Homo sapiens clone RP11-206N2, WORKING DRAFT SEQUENCE, 37 unordered
            pieces.
ACCESSION   AC068682
VERSION     AC068682.3  GI:10280868
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 184536)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-206N2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184536)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 23, 2000 this sequence version replaced gi:8247861.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5802
                Center clone name: 206_N_2
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 167659 bases at least Q40
                Consensus quality: 175736 bases at least Q30
                Consensus quality: 178914 bases at least Q20
                Insert size: 154000; agarose-fp
                Insert size: 180936; sum-of-contigs
                Quality coverage: 5.0 in Q20 bases; agarose-fp
                Quality coverage: 4.3 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1154: contig of 1154 bp in length
            *    1155     1254: gap of 100 bp
            *    1255     2766: contig of 1512 bp in length
            *    2767     2866: gap of 100 bp
            *    2867     4073: contig of 1207 bp in length
            *    4074     4173: gap of 100 bp
            *    4174     5191: contig of 1018 bp in length
            *    5192     5291: gap of 100 bp
            *    5292     7387: contig of 2096 bp in length
            *    7388     7487: gap of 100 bp
            *    7488     9087: contig of 1600 bp in length
            *    9088     9187: gap of 100 bp
            *    9188    10361: contig of 1174 bp in length
            *   10362    10461: gap of 100 bp
            *   10462    12065: contig of 1604 bp in length
            *   12066    12165: gap of 100 bp
            *   12166    13498: contig of 1333 bp in length
            *   13499    13598: gap of 100 bp
            *   13599    15085: contig of 1487 bp in length
            *   15086    15185: gap of 100 bp
            *   15186    17249: contig of 2064 bp in length
            *   17250    17349: gap of 100 bp
            *   17350    19243: contig of 1894 bp in length
            *   19244    19343: gap of 100 bp
            *   19344    21127: contig of 1784 bp in length
            *   21128    21227: gap of 100 bp
            *   21228    23345: contig of 2118 bp in length
            *   23346    23445: gap of 100 bp
            *   23446    25965: contig of 2520 bp in length
            *   25966    26065: gap of 100 bp
            *   26066    28195: contig of 2130 bp in length
            *   28196    28295: gap of 100 bp
            *   28296    32216: contig of 3921 bp in length
            *   32217    32316: gap of 100 bp
            *   32317    36954: contig of 4638 bp in length
            *   36955    37054: gap of 100 bp
            *   37055    40348: contig of 3294 bp in length
            *   40349    40448: gap of 100 bp
            *   40449    43556: contig of 3108 bp in length
            *   43557    43656: gap of 100 bp
            *   43657    47926: contig of 4270 bp in length
            *   47927    48026: gap of 100 bp
            *   48027    54079: contig of 6053 bp in length
            *   54080    54179: gap of 100 bp
            *   54180    58962: contig of 4783 bp in length
            *   58963    59062: gap of 100 bp
            *   59063    63739: contig of 4677 bp in length
            *   63740    63839: gap of 100 bp
            *   63840    71551: contig of 7712 bp in length
            *   71552    71651: gap of 100 bp
            *   71652    78218: contig of 6567 bp in length
            *   78219    78318: gap of 100 bp
            *   78319    84982: contig of 6664 bp in length
            *   84983    85082: gap of 100 bp
            *   85083    91930: contig of 6848 bp in length
            *   91931    92030: gap of 100 bp
            *   92031    98922: contig of 6892 bp in length
            *   98923    99022: gap of 100 bp
            *   99023   106496: contig of 7474 bp in length
            *  106497   106596: gap of 100 bp
            *  106597   114950: contig of 8354 bp in length
            *  114951   115050: gap of 100 bp
            *  115051   124353: contig of 9303 bp in length
            *  124354   124453: gap of 100 bp
            *  124454   132639: contig of 8186 bp in length
            *  132640   132739: gap of 100 bp
            *  132740   141572: contig of 8833 bp in length
            *  141573   141672: gap of 100 bp
            *  141673   151302: contig of 9630 bp in length
            *  151303   151402: gap of 100 bp
```

```
          *   151403   161767: contig of 10365 bp in length
          *   161768   161867: gap of 100 bp
          *   161868   184536: contig of 22669 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184536
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-206N2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1154
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             1155. .1254
                     /estimated_length=100
     misc_feature    1255. .2766
                     /note="assembly_fragment"
     gap             2767. .2866
                     /estimated_length=100
     misc_feature    2867. .4073
                     /note="assembly_fragment"
     gap             4074. .4173
                     /estimated_length=100
     misc_feature    4174. .5191
                     /note="assembly_fragment"
     gap             5192. .5291
                     /estimated_length=100
     misc_feature    5292. .7387
                     /note="assembly_fragment"
     gap             7388. .7487
                     /estimated_length=100
     misc_feature    7488. .9087
                     /note="assembly_fragment"
     gap             9088. .9187
                     /estimated_length=100
     misc_feature    9188. .10361
                     /note="assembly_fragment"
     gap             10362. .10461
                     /estimated_length=100
     misc_feature    10462. .12065
                     /note="assembly_fragment"
     gap             12066. .12165
                     /estimated_length=100
     misc_feature    12166. .13498
                     /note="assembly_fragment"
     gap             13499. .13598
                     /estimated_length=100
     misc_feature    13599. .15085
                     /note="assembly_fragment"
     gap             15086. .15185
                     /estimated_length=100
     misc_feature    15186. .17249
                     /note="assembly_fragment"
     gap             17250. .17349
                     /estimated_length=100
     misc_feature    17350. .19243
                     /note="assembly_fragment"

  Query Match             39.7%;  Score 793.6;  DB 14;  Length 184536;
  Best Local Similarity   68.2%;  Pred. No. 3.3e-200;
  Matches 1339;  Conservative    0;  Mismatches  444;  Indels  179;  Gaps   10;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||  ||||||||  ||||||||||||  ||||||||||||
Db     138326 AAAGGCTTCTGAAATCAGACAACACCTTTCAAGCTCTTATACCAACTTCTGGAGTTGGGC 138385

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||  || ||||||||||||  || || | |||||
Db     138386 AACATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTGTTACTCACCTTTGG 138445

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||  |||||||||||| || ||||||||||||||||| | |||  ||||||||||||

Db     138446 GCCCTATATTTTTAACCTTTTTGTCAAATTTGTTTCCTCTGGAATCAAGGCCATCAAGCC 138505

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              |||||||  ||||||||  ||  |||  |||||||||  |||||  ||||||||||  | |||||||
Db     138506 ACAGATGGTCTTACAAGTGGAACACCAAATGAGTTCAACCAACAACTTCTACCGAGGACC 138565

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||  ||||  |||||||||  ||  ||||||||||||  ||||||||||||||||  |||
Db     138566 CCTGGACTGACCTGCTGGCCCTTCCACTGGCCTAAAGAGTTCCCCTCTGGAGGACATTAC 138625

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG--C 358
              ||||||||||  |||||||  |||  ||||||||||||||||||||  ||||||||||  |
Db     138626 AACTGCAGGGCTCCTTCTTTGCCCTTATCCAGCAGGAAGTAGCTAAAGCGGTCATCGGCC 138685

Qy        359 CAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGC 418
              ||  ||||||||||||  |||||  ||||||||  ||||||||||||||||||||||||  ||
Db     138686 AAATTCCCAACAGCAGTTGGGGCATCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCGGC 138745

Qy        419 TGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              ||||||||||||   ||||||||||||||||  ||||  ||||||||||||||||||||||
Db     138746 TGGGCTTCTGGGCTGGGTGGGGACTTGGAGAA-TTTTCTGTCTAGCTAAAGGATTGTAAA 138804

Qy        479 TGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATT---------------------- 514
              ||||||||||||  ||||||||||||||||| | ||
Db     138805 TGCACCAATCAGAGCTCTGTGTCTAGCTAAAAGGTTTGTAAACACACCAATCAGCACTCT 138864

Qy        515 -GTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTC 573
               |||||  ||||||||||  ||||||||||||||||||||||||||||||||||||  | |
Db     138865 GTAAAATGGACCAATCAGCTCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCC 138924

Qy        574 AAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTT 633
              ||||||||||  |||||  ||||||||||  ||  ||||||||||||||  |||  |||||||||
Db     138925 AAATAAGGGAATAAAAGCTGGCCACCCAAGTCAGCAGTGGCAACCCGCTCAGGTCCCCTT 138984

Qy        634 CCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCT 693
              |||  ||||||||||||||||||||  ||||||  ||||||||||||||||||||||  |||
Db     138985 CCATGCTGTGGAAGCTTTGTTCTTTCACTCTTCGCAATAAATCTTGCTGCTGCTCACTCT 139044

Qy        694 TTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTC 753
              |||  ||||  |||||  |||  |||||  |||  |||||  ||  |  ||  ||||||  ||||||  ||
Db     139045 TTGGGTCCGCACTAACTTCATGAGTTGTCACACTTACCGTGAAGGTCTGCAGCTTCACTC 139104

Qy        754 CTGAAGTCA---------------------------------------------------- 762
              ||||||  ||
Db     139105 CTGAAGCCACTGAGACCATGGACCCACCAGGAGGAAAAAGCAACTCCAGACGTGCCACCT 139164

Qy        763 ----------------------------------------------------ACAGACC 769
                                                                   |||||
Db     139165 TTAAGAGCTGTAACACTCACCGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACC 139224

Qy        770 ACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACT 829
              ||  ||||||||||||||||||||  ||||||  ||  |   |  ||||||||||||||  ||
Db     139225 ACAAACCCACTGGAAGGAACAAACAACTCCTGACGCACCACCTTTAAGAGCTGTAACCCT 139284

Qy        830 CACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGG 889
              |||  |  ||||  ||||  ||||||||||||||||||||  ||||||||  ||||||||||||||
Db     139285 CACAGTGAAGGTCTGAAGCTTCACTCCTGAAGTCAGCGAGACCACGAACCCACCAGAAGG 139344

Qy        890 AAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCA 949
              |||||||||  |||  |||  |||||  ||||||||||||||||||  |    |||||||||  |
Db     139345 AAGAAACTCCGGATACATCTGAACATCTGAAGGAACAAACTCCGG----ACCATCTTTAA 139400

Qy        950 GAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGA 1009
              ||  ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db     139401 GAACTGTAACACTCACCGCAAGGGTCTGTGGCTTTGTTCTTGAAGTCAGCAAGACCAAGA 139460

Qy       1010 ACCCACCGGAAGGAACAAATTCCAGACACAGTAGG------------------------A 1045
              |  ||||  ||||||||  ||||||  ||||||  |||                     |
Db     139461 ATCCACTGGAAGGAATCAATTCCGGACACATAAGGCCGGGCACGGTGGCTCACGCCTGCA 139520

Qy       1046 AATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCA 1105
              |  |      ||||    ||  |   |  ||   | | ||     ||  ||   |
Db     139521 ATCCCAGCACTTTGGGAGGTCGAGGCAGGTGGATCACGAGGTTAGGAGATCAAGACCATC 139580
```

```
Qy       1106 GCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGC 1165
Db     139581 CTGGCTAACACAGTGAAACCCCGTCTCTACTGAAAATACAAAAACATTAGCCGAGCATGG 139640

Qy       1166 TTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATA 1225
Db     139641 TGGCAGTCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGCGTAAACCTG 139700

Qy       1226 ATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTT------GCATCTGAGACATGT 1279
Db     139701 GGAGGCGGAGCTTGCAGTGAGCTGAGATTGTGCCACTGTACTCCAGCCTGGGTGACAGAG 139760

Qy       1280 AGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCGATGGCTCATGCCTGTAAT 1339
Db     139761 TGAGACTCCATCTCAAAAAAAAAAAAAAAAAATGCTGGGCGTGATGGCTCACACCTGTAAT 139820

Qy       1340 CCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCAGGCGTTCGAGACCAGCCT 1399
Db     139821 CCCAGCACTTTGGGAGGCCGAGGCGGGCAGACCACCAGGTCAGGAGATCAAGACCATCCT 139880

Qy       1400 GGCCAACATGGTGAAACCCCGTCTCTACTAAAAA------TACAAAAATTAGCCTGGTGT 1453
Db     139881 GGCCAACATGGTGAAACCCCCTCTCCACTAAAAAAAATACAAAAAAAATTAGCCAGGCAT 139940

Qy       1454 GGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAAC 1512
Db     139941 GGGGGTGAGCGTCTGTAGTCCTGGCTACTCCGGAGGCTAAGGCAGGAGAATGGCGTGAAC 140000

Qy       1513 TTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCATCCTGGGTGACA 1572
Db     140001 CCGGGAGGTGGAGCTTGCAATGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCAACA 140060

Qy       1573 GAGCGAGACTCTGTCTCAAA------AAAAAAAAAAAAAAAAGGAAAGGAAAATATAATC 1626
Db     140061 GAGCGAGACTCCGTCTCAAAGAAAATAAATAAATAAAAAATAAAAATAAATAAATAATA 140120

Qy       1627 AAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACACTAAGT 1686
Db     140121 AAAAAGATAAAGTGGTAGGCTTAAGCCCAAACATATCAATAGTTACATTAAATGTAAATG 140180

Qy       1687 GTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGCTGTTAT 1746
Db     140181 GTTTAACTATAAAAATTAAAATACAGTAACTAGCATAATGGATTAAAAACATGCCCTAAT 140240

Qy       1747 TATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGT 1788
Db     140241 TATATGTTCTCTACAAGAAACTCACATCAAATACAATAACGT 140282
```

RESULT 14
HSU116E9

LOCUS	HSU116E9 25158 bp DNA linear PRI 18-MAY-2005
DEFINITION	Human DNA sequence from clone LL0XNC01-116E9 on chromosome Xq22, complete sequence.
ACCESSION	Z95333
VERSION	Z95333.2 GI:23304713
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 25158)
AUTHORS	Grafham,D.
TITLE	Direct Submission
JOURNAL	Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT	On Sep 23, 2002 this sequence version replaced gi:2213445. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX LL0XNC01-116E9 is from the Lawrence Livermore National Laboratory flow-sorted X chromosome cosmid library LL0XNC01 VECTOR: lawrist16 ---------- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: vega@sanger.ac.uk ---------- This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES Location/Qualifiers

source	1. .25158
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="RZPD:LLNLc110E09116"
	/db_xref="taxon:9606"
	/map="q22"
	/clone="LL0XNC01-116E9"
	/clone_lib="LL0XNC01"

ORIGIN

Query Match 39.2%; Score 785.2; DB 8; Length 25158;
Best Local Similarity 87.2%; Pred. No. 6e-198;
Matches 919; Conservative 0; Mismatches 128; Indels 7; Gaps 5;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
Db       9443 AAAGGCTTCTGAAATCACACAACACCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 9502

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
Db       9503 AACATGGCTTCTCCCCTTTCTCGGTCCCATTGCAGCCATCTTGCTATTACTCGCCTTCAG 9562

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
Db       9563 GCTGTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGAATTGAGGCCGTCAAGCT 9622

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
Db       9623 ACAGATGGTCTTACAAATGGGACCCCAAATGAGCTCAACTAACAACTTCTGCCAAGGACC 9682

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
Db       9683 CCTGGACCAACCTGCTGGCCCTTTCACTGGCCTTAAGAGTTCCCCTCTGGAGGGCACTAC 9742

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
Db       9743 AACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGCAGTCATCACCC 9802

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
Db       9803 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGAGATTGAGAGTTCAAGGCAGCT 9862

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
Db       9863 GGGCTTCTGGGTTGGGTGGAGACTTGGAGAGCTTTTCTGTCTAGCTAAAGGATTTTAAAC 9922

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT 539
Db       9923 ACACCAATCAGTGCTC--TATCTAGCTAAAGGTTTGTAAATGCACCAAGCAGCACTCTGT 9980
```

```
Qy        540 AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACC 599
              ||||||||||||||||||||||||||| |||| || |||||||| ||||| |||||||||
Db       9981 AAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGGGAATAAAAGCTGGCCACC 10040

Qy        600 CGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTT 659
              |||||||||| |||||||||| |||| ||||||| ||||||||||||||||||||||
Db      10041 TGAGCCAGCAGCAGCAACCCACTTTGGTCACCTTCCATGCTGTGGAAGCTTTGTTCTTTT 10100

Qy        660 GCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCT 719
              ||||||||||||||||||||||||| |||| |||||| ||||||||||||||||||||||
Db      10101 GCTCTTCACAATAAATCTTGCTGCTCCTCACTCTTTGGGTCCACACTACCTTTATGAGCT 10160

Qy        720 GTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCACGAACCCA 778
              |||||||||||||||| |||||| |||||| |||||||||| ||||||| ||||||
Db      10161 GTAACACTCACTGCGAAGGTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCA 10220

Qy        779 CTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAA 838
              |||  ||||| ||| |||||| |||| | ||||||||||||||||||| | |||||||
Db      10221 CTGAGAGGAATAAACAACTCCGGATGCCCCACCTTTAAGAGCTGTAACACACCCTGCGAA 10280

Qy        839 GCTCTGCAGCTTCACTCCTGAAGTC-AGTGAGACCACAAACCCACCAGAAGGAAGAAACT 897
              |||||  |||||||||||||||| || || || ||||  ||||||||||||||||||||
Db      10281 AGTCTGCGGCTTCACTCCTGAAATCAAGAGATACCATGAACCCACCAGAAGGAAGAAA-- 10338

Qy        898 CTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTA 957
              ||||||||| ||||| |||||||||||  |||||  |||||||||||||| ||| |||||
Db      10339 CTGGACACATCTGAACATCTGAAGGAATGAACTCTGGACACACCATCTTTAAGAACTGTA 10398

Qy        958 ACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCG 1017
              |||||||||   || |||    ||||| |||| ||||||||| |||||||||||||| |
Db      10399 ACACTCACCATGAGAGTCCACAGCTTCGTTCTCGAAGTCAGCGAGACCAAGAACCCACTG 10458

Qy       1018 GAAGGAACAAATTCCAGACACAGTAGGAAATCTG 1051
              |||||||| |||||  |||||||||  | |  ||
Db      10459 GAAGGAACGAATTCTGGACACAGTAATATAGATG 10492
```

```
RESULT 15
AC092017
LOCUS       AC092017               193041 bp    DNA     linear   PRI 20-NOV-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-543B16, complete sequence.
ACCESSION   AC092017 AL356972
VERSION     AC092017.2  GI:17017614
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 193041)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193041)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 193041)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Nov 20, 2001 this sequence version replaced gi:14423605.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: SC
            -------------- Project Information
                Center project name: chr-1
                Center clone name: RP11-543B16 (sc0173)
            ------------ Summary Statistics
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator ET; 88% of reads
                Chemistry: Dye-terminator Big Dye; 12% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 192648 bases at least Q40
                Consensus quality: 193013 bases at least Q30
                Consensus quality: 193041 bases at least Q20
                Insert size: 200871; 9.7% error; agarose-fp
                Insert size: 193041; sum-of-contigs
                Quality coverage: 6.5x in Q20 bases; agarose-fp
                Quality coverage: 6.7x in Q20 bases; sum-of-contigs
            --------------.
            Overlapping Sequences:
                5': Mapping in progress
                3': Mapping in progress
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
```

BglII		EcoRI		HindIII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
3118	3023	8696	8671	2515	2428
2067	2109	6	<800	6382	6499
9978	9634	62	<800	512	<800
565	<800	3928	3905	449	<800
2339	2376	12181	11850	4775	4829
9773	9634	13497	13015	5390	5382
1140	1152	1482	1462	485	<800
1446	1435	1927	1935	7066	6975
758	764	4640	4606	1525	1475

16631	16312	8720	8671	775	771
6773	6774	1510	1462	887	891
376	<800	7700	7817	541	<800
3343	3344	3100	3152	6795	6975
8645	8582	723	730	5463	5382
2983	3023	1151	1071	2991	2987
5895	5784	15253	14879	1731	1769
1622	1606	7131	7188	4165	4095
5103	5075	2118	2064	3185	3214
6740	6774	1959	1935	664	<800
1905	1923	5491	5504	14342	14386
1637	1606	16	<800	8040	8002
1013	998	1090	1131	155	<800
526	<800	677	<800	1404	1338
11090	10902	57	<800	3025	3214
1175	1152	6052	6084	10582	10416
5066	5075	5140	5037	6178	6022
5214	5075	3074	3152	10309	10416
1475	1435	981	979	4191	4095
2313	2376	2669	2714	7361	7451
2340	2376	2223	2298	2667	2698
5769	5784	1309	1286	4804	4829
882	892	5371	5504	4832	4829
4356	4356	147	<800	3587	3548
6696	6774	15923	15755	3850	3829
4468	4356	1505	1462	2406	2428
3929	3914	8557	8671	9365	9375
7412	7695	3819	3905	734	771
15854	16312	1655	1642	11752	11688
6113	6108	6262	6354	310	<800
266	<800	23009	23575	5989	6022
102	<800	712	<800	102	<800
4345	4356	5827	6084	75	<800
603	<800	709	<800	1091	1079
1142	1152	1067	1071	4518	4481
6905	7396	2614	2714	4943	4829

2368	2376	6296	6311
3891	3914	4286	4333
34	<800	5345	5382
2135	2109	2905	2987
1421	1435		

```
FEATURES             Location/Qualifiers
     source          1. .193041
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-543B16"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    55137. .55207
                     /note="Single subclone region"
ORIGIN
```

Query Match 39.1%; Score 782; DB 8; Length 193041;
Best Local Similarity 86.9%; Pred. No. 4.1e-197;
Matches 928; Conservative 0; Mismatches 110; Indels 30; Gaps 5;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||  |||| ||||||||||||||||||||||||||||||||||  ||
Db     171360 AAAGGCTTCTGAAATCAAACAACGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGTGC 171419

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||  ||||||||||| |||||  ||  ||  | ||  ||
Db     171420 AACATGGCTTCTCCCCTTTCTAGGTCCCATGACAGCCATCCTGCTATTACTCACCTTCGG 171479

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||| | || ||||||||||  |  ||||||||||||||||||  || ||||||||||||||
Db     171480 GCCATATA-TTTTAACCTCCTTATCAAATTTGTTTCCTCTAGAATTGAGGCCATCAAGCC 171538

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              |||||||  ||||||||||| |||||||||  |||||||||||||||||||  |  ||||||
Db     171539 ACAGATGGTCTTACAAATGGAACCCCAAACGAGCTCAACTAACAACTTCTACCAAGGACC 171598

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||  ||||||||||||  ||||||||||||||||||||||||||||||||
Db     171599 CCTGGACCGACCCACTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 171658

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCA-TCGCC 359
              |||||||||| ||||||||||||| | ||||||||||||||  |||||||  | |||
Db     171659 AACTGCAGGGCTCCTTCTTCACCCCTGTTCAGCAGGAAGTAGCTAGGGCGGTCATTGGCC 171718

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              ||  ||||||||||||  ||||||||||||||||  || ||||||||||||||||||  |||
Db     171719 AATTCCCAACAGCAGTTGGGGTGTCCTGTTT--AGAGGGGATTGAGAGGTGAAGCTGGCT 171776

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              ||||||||||||  ||||||||||||||||  ||||||||  ||||||||||||  ||||||||
Db     171777 GGGCTTCTGGGTCGGGTGGGGACTTGGGGAACTTTTCTGTCTAGCTAAAGTATTGTAAAC 171836

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCA------ 533
              ||||||||||  ||||||||||||||||||||  ||||||  ||||||||||||
Db     171837 ACACCAATCAGTGCTCTGTGTCTAGCTAAAGGGTTGTAAGCGCACCAATCAGCACTCTGT 171896

Qy        534 -------------------CTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCA 574
                                  |||||||||||||||||||||||||||||||||||| |||| ||
Db     171897 AAAAACGGACCAATTGGCTCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCA 171956

Qy        575 AATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTC 634
              |||||||||  |||||  |||||||||  ||||||  |||||||  |||||||  || |||
Db     171957 AATAAGGGAATAAAAGCTGGCCACCTGAGCCAGCAGCAGCAACCTGCTCGGGTGCCTTTC 172016

Qy        635 CACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTT 694
              |||||  |||||||||||||||||  ||||||||||||||||||||||||||||||  | ||
Db     172017 CACACCGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCTGCTCACTGTT 172076
```

```
Qy        695 TGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCC 754
              || |||  |||||||||||||||||||||||| |||| | | |||||   |||||| |||
Db     172077 TGGGTCTGCACTACCTTTATGAGCTGTAACACGCACTTCAAAGGTCTACAGCTTCACTCC 172136

Qy        755 TGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813
              |||||||| | ||||||| ||||||| ||||||||||||  |||   || ||||  ||||
Db     172137 TGAAGTCAGCGAGACCACAAACCCACAGGAAGGAACAAACGACTATGGACGTGCCACCTT 172196

Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873
              |||||||||||||||||||| |||| ||||  |||||||||||||||||||  ||||||
Db     172197 TAAGAGCTGTAACACTCACTGTGAAGGTCTGTGGCTTCACTCCTGAAGTCAGCAAGACCA 172256

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933
              | |||||||||||||||||||||||||||||| ||||| ||||||||||||||||||||
Db     172257 CGAACCCACCAGAAGGAAGAAACTCTGGACACATCTGAACATCTGAAGGAACAAACTCCA 172316

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
              ||||||| |||||| ||| |||||||||||||| | ||||||  ||||||||||||||||
Db     172317 GACACACGATCTTTAAGAACTGTAACACTCACCACGAGGGTCCATGGCTTCATTCTTGAA 172376

Qy        994 GTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGT 1041
              |||||| |||||||||||||| ||||||||| |||||| ||||||||
Db     172377 GTCAGCGAGACCAAGAACCCACAGGAAGGAACCAATTCCGGACACAGT 172424
```

```
RESULT 16
AC011098/c
LOCUS       AC011098              162397 bp    DNA     linear   HTG 16-JUN-2000
DEFINITION  Homo sapiens clone RP11-2L1, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   AC011098
VERSION     AC011098.3  GI:8569146
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 162397)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-2L1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162397)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:8096843.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1549
                Center clone name: 2_L_1
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 150661 bases at least Q40
                Consensus quality: 155772 bases at least Q30
                Consensus quality: 158196 bases at least Q20
                Insert size: 159000; agarose-fp
                Insert size: 160997; sum-of-contigs
                Quality coverage: 4.4 in Q20 bases; agarose-fp
                Quality coverage: 4.3 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1728: contig of 1728 bp in length
            *     1729     1828: gap of 100 bp
            *     1829     3274: contig of 1446 bp in length
            *     3275     3374: gap of 100 bp
            *     3375     4354: contig of 980 bp in length
            *     4355     4454: gap of 100 bp
            *     4455     7334: contig of 2880 bp in length
            *     7335     7434: gap of 100 bp
            *     7435     9828: contig of 2394 bp in length
            *     9829     9928: gap of 100 bp
            *     9929    14825: contig of 4897 bp in length
            *    14826    14925: gap of 100 bp
            *    14926    21182: contig of 6257 bp in length
            *    21183    21282: gap of 100 bp
            *    21283    29127: contig of 7845 bp in length
            *    29128    29227: gap of 100 bp
            *    29228    38966: contig of 9739 bp in length
            *    38967    39066: gap of 100 bp
            *    39067    50624: contig of 11558 bp in length
            *    50625    50724: gap of 100 bp
            *    50725    67922: contig of 17198 bp in length
            *    67923    68022: gap of 100 bp
            *    68023    83588: contig of 15566 bp in length
            *    83589    83688: gap of 100 bp
            *    83689    98121: contig of 14433 bp in length
            *    98122    98221: gap of 100 bp
            *    98222   128190: contig of 29969 bp in length
            *   128191   128290: gap of 100 bp
            *   128291   162397: contig of 34107 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162397
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-2L1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1728
                     /note="assembly_fragment"
     gap             1729. .1828
                     /estimated_length=100
     misc_feature    1829. .3274
                     /note="assembly_fragment"
     gap             3275. .3374
                     /estimated_length=100
     misc_feature    3375. .4354
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     gap             4355. .4454
                     /estimated_length=100
     misc_feature    4455. .7334
                     /note="assembly_fragment"
     gap             7335. .7434
```

```
                     /estimated_length=100
     misc_feature    7435. .9828
                     /note="assembly_fragment"
     gap             9829. .9928
                     /estimated_length=100
     misc_feature    9929. .14825
                     /note="assembly_fragment"
     gap             14826. .14925
                     /estimated_length=100
     misc_feature    14926. .21182
                     /note="assembly_fragment"
     gap             21183. .21282
                     /estimated_length=100
     misc_feature    21283. .29127
                     /note="assembly_fragment"
     gap             29128. .29227
                     /estimated_length=100
     misc_feature    29228. .38966
                     /note="assembly_fragment"
     gap             38967. .39066
                     /estimated_length=100
     misc_feature    39067. .50624
                     /note="assembly_fragment"
     gap             50625. .50724
                     /estimated_length=100
     misc_feature    50725. .67922
                     /note="assembly_fragment"
     gap             67923. .68022
                     /estimated_length=100
     misc_feature    68023. .83588
                     /note="assembly_fragment"
     gap             83589. .83688
                     /estimated_length=100
     misc_feature    83689. .98121
                     /note="assembly_fragment"
     gap             98122. .98221
                     /estimated_length=100
     misc_feature    98222. .128190
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             128191. .128290
                     /estimated_length=100
     misc_feature    128291. .162397
                     /note="assembly_fragment"
ORIGIN

  Query Match             38.9%;  Score 779;  DB 14;  Length 162397;
  Best Local Similarity   86.9%;  Pred. No. 2.6e-196;
  Matches  917;  Conservative    0;  Mismatches  125;  Indels   13;  Gaps    5;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||| ||||||||||||||||||  ||||||||||||||||||||||| |||| |||
Db     145936 AAAGTCTTCTGAAATCAGACAATGCTCTTCAAACTCTTATACCAACCTCTGAAGTTAGGC 145877

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||  ||||||||||| ||| || ||   ||  |
Db     145876 AACATGGCTTCTCCCCTTTCTAGGTCCCATGACAGCCATCTTACTATTACTCATCTTCAG 145817

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||||||||||| | ||||||||||
Db     145816 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATCAATGCCATCAAGCT 145757

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||| |||||||||| | |||||||
Db     145756 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAATTAACAACTTCTACCGAGGACC 145697

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||| ||| |||||||||| || |||||||||||| ||||||||||||||||||||
Db     145696 CCTGGACCAACCTGCTGGCCCTTCCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 145637

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCC- 359
              |||||||||||||||||  |||| ||||||||||||||||||| ||| |||| | ||

Db     145636 ATCTGCAGGGCCCCTTCTTTGCCCC-ATCCAGCAGGAAGTAGCTAGAGCAGTCACCACCT 145578

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||||||| |||||||||||||||||||||||||||||
Db     145577 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCAGCT 145518

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||| |||||||||||||||||||| |||||||||| ||||||||||||
Db     145517 GGGCTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAGAGGATTGTAAAT 145458

Qy        480 GCACCAATCAGCACTCTGTGTCTAGC--------TAAAGGATTGTAAATGCACCAATCAG 531
              |||||||||||| ||| |      |         |   | |||   |||||||||
Db     145457 GCACCAATCAGCGCTCAATAAAACACACCAATCAGCACTCTGTAAAAACAGACCAATCAG 145398

Qy        532 CACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAAC 591
              |||||||||||||||||||||||||||||||||| |||| |||||||||||| ||||| |
Db     145397 CACTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAGGGAATAAAAGC 145338

Qy        592 TGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTT 651
              |||||||| | |||||||  ||||| | || ||||||||||||| ||||||||||||
Db     145337 TGGCCACCCAACCCAGCAGCAGCAACACGCTTGGGTCCCCTTCCACGCTGTGGAAGCTTT 145278

Qy        652 GTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTT 711
              |||||||||||||||||||||||||||||||||| ||||   |||||| ||  ||
Db     145277 GTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTCAGGTCCACGCT--CTC 145220

Qy        712 TATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCA 770
              |||||||||||||||||||||||||||||||  |||||||||||||||||| | ||||||
Db     145219 TATGAGCTGTAACACTCACTGCGAGGGTCTGCAGCTTCATTCCTGAAGTCAGCGAGACCA 145160

Qy        771 CGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTC 830
              | ||||||| || |||||||| |||||  || | ||  |||||||||||||||||| ||
Db     145159 CAAACCCACAGGGAGGAACAAACAACTCTGGACGCGCCACCTTTAAGAGCTGTAACATTC 145100

Qy        831 ACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGA 890
              ||||| ||  ||||| ||||||||||||||||||  ||||||  |||||||  ||||||
Db     145099 ACTGCAAAAGTCTGCGGCTTCACTCCTGAAGTCAGCAAGACCATGAACCCACAGGAAGGA 145040

Qy        891 AGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAG 950
              ||||||||  ||||||  ||||| |||||||||||||||||  |||||||||||||| |
Db     145039 AGAAACTCCAGACACATCTGAACATCTGAAGGAACAAACTCTGGACACACCATCTTTAAC 144980

Qy        951 AGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAA 1010
              | |||||||||||||  ||||||  || ||||||||||||||||||||| ||||||||||
Db     144979 AACTGTAACACTCACTGCAAGGATCCGTGGCTTCATTCTTGAAGTCAGCGAGACCAAGAA 144920

Qy       1011 CCCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
              |||||||||||||||  |||||  ||||||  || ||
Db     144919 CCCACCGGAAGGAACCAATTCTGGACACACTATGA 144885


RESULT 17
AC011328/c
LOCUS       AC011328              216031 bp    DNA     linear   PRI 09-APR-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-295O23, complete sequence.
ACCESSION   AC011328
VERSION     AC011328.11  GI:19744957
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 216031)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216031)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1999) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
```

```
REFERENCE   3  (bases 1 to 216031)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   4  (bases 1 to 216031)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   5  (bases 1 to 216031)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   6  (bases 1 to 216031)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Mar 27, 2002 this sequence version replaced gi:15021986.
            -------------- Genome Center
                Center: Genome Therapeutics Corporation
                Center code: GTC
                Web site: http://www.genomecorp.com/
                Contact: gtc-seqcenter@genomecorp.com
            -------------- Project Information
                Center project name: hg031

                IMPORTANT: This sequence is the entire
                insert of clone RP11-295O23.
                The true right end of clone RP11-506P9 is at
                2000 in this sequence.
            -------------- Summary Statistics
                Sequencing vector: N/A
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 990315
            --------------.
FEATURES             Location/Qualifiers
     source          1. .216031
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-295O23"
                     /clone_lib="RPCI-11"
ORIGIN

  Query Match             38.9%;  Score 779;  DB 8;  Length 216031;
  Best Local Similarity   86.9%;  Pred. No. 2.6e-196;
  Matches  917;  Conservative    0;  Mismatches  125;  Indels   13;  Gaps    5;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||| ||||||||||||||||||||  |||||||||||||||||||||| |||| |||
Db     187903 AAAGTCTTCTGAAATCAGACAATGCTCTTCAAACTCTTATACCAACCTCTGAAGTTAGGC 187844

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||  |||||||||||| ||| || ||   ||  |
Db     187843 AACATGGCTTCTCCCCTTTCTAGGTCCCATGACAGCCATCTTACTATTACTCATCTTCAG 187784

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | ||||||||||||||||||||| | ||||||||||
Db     187783 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATCAATGCCATCAAGCT 187724

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||| |||||||||| | |||||||
Db     187723 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAATTAACAACTTCTACCGAGGACC 187664

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||| ||| |||||||||| || |||||||||||| ||||||||||||||||||||
Db     187663 CCTGGACCAACCTGCTGGCCCTTCCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 187604

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCC- 359
              |||||||||||||||||  |||| ||||||||||||||||||| ||| |||| | ||
Db     187603 ATCTGCAGGGCCCCTTCTTTGCCCC-ATCCAGCAGGAAGTAGCTAGAGCAGTCACCACCT 187545

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||||||| |||||||||||||||||||||||||||||
Db     187544 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCAGCT 187485

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||||| |||||||||||||||||||||| |||||||||| ||||||||||||
Db     187484 GGGCTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAGAGGATTGTAAAT 187425

Qy        480 GCACCAATCAGCACTCTGTGTCTAGC--------TAAAGGATTGTAAATGCACCAATCAG 531
              |||||||||||| ||| |       |           |      | |||   |||||||||
Db     187424 GCACCAATCAGCGCTCAATAAAACACACCAATCAGCACTCTGTAAAAACAGACCAATCAG 187365

Qy        532 CACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAAC 591
              |||||||||||||||||||||||||||||||||||| |||| |||||||||| ||||| |
Db     187364 CACTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAGGGAATAAAAGC 187305

Qy        592 TGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTT 651
              ||||||||| | ||||||| ||||| | || |||||||||||||| ||||||||||||
Db     187304 TGGCCACCCAACCCAGCAGCAGCAACACGCTTGGGTCCCCTTCCACGCTGTGGAAGCTTT 187245

Qy        652 GTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTT 711
              |||||||||||||||||||||||||||||||||||| ||||  |||||| ||  ||
Db     187244 GTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTCAGGTCCACGCT--CTC 187187

Qy        712 TATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCA 770
              |||||||||||||||||||||||||||||  |||||||||||||||||| | ||||||
Db     187186 TATGAGCTGTAACACTCACTGCGAGGGTCTGCAGCTTCATTCCTGAAGTCAGCGAGACCA 187127

Qy        771 CGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTC 830
              | ||||||| || ||||||||| |||||  || | ||  |||||||||||||||||| ||
Db     187126 CAAACCCACAGGGAGGAACAAACAACTCTGGACGCGCCACCTTTAAGAGCTGTAACATTC 187067

Qy        831 ACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGA 890
              ||||| ||  ||||| |||||||||||||||||||  ||||||  |||||||  ||||||
Db     187066 ACTGCAAAAGTCTGCGGCTTCACTCCTGAAGTCAGCAAGACCATGAACCCACAGGAAGGA 187007

Qy        891 AGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAG 950
              ||||||||  ||||||  ||||| |||||||||||||||||  |||||||||||||| |
Db     187006 AGAAACTCCAGACACATCTGAACATCTGAAGGAACAAACTCTGGACACACCATCTTTAAC 186947

Qy        951 AGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAA 1010
              | |||||||||||||| |||||| || |||||||||||||||||||||| ||||||||||
Db     186946 AACTGTAACACTCACTGCAAGGATCCGTGGCTTCATTCTTGAAGTCAGCGAGACCAAGAA 186887

Qy       1011 CCCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
              ||||||||||||||| |||||  |||||| || ||
Db     186886 CCCACCGGAAGGAACCAATTCTGGACACACTATGA 186852

RESULT 18
AL390836/c
LOCUS       AL390836              169482 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP13-179C6 on chromosome X, complete
            sequence.
ACCESSION   AL390836
VERSION     AL390836.12  GI:11225794
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 169482)
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 19, 2000 this sequence version replaced gi:10277989.
            The following abbreviations are used to associate primary accession
```

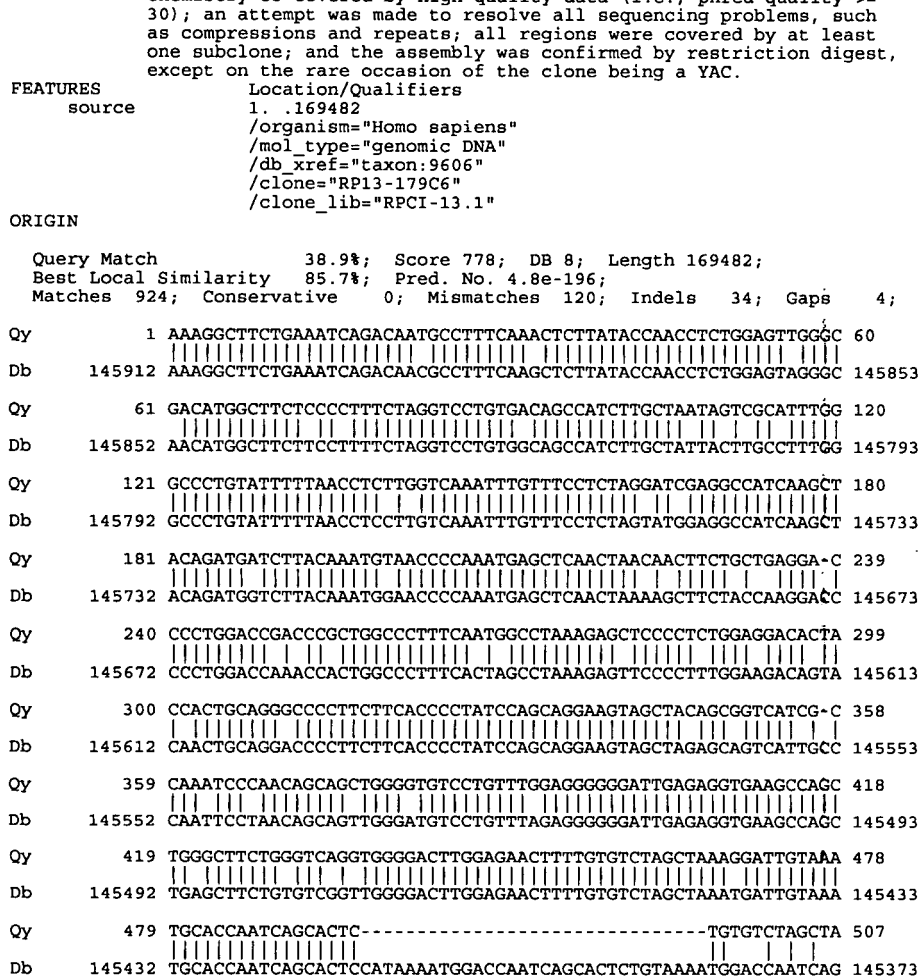
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP13-179C6 is from the library RPCI-13.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vega@sanger.ac.uk

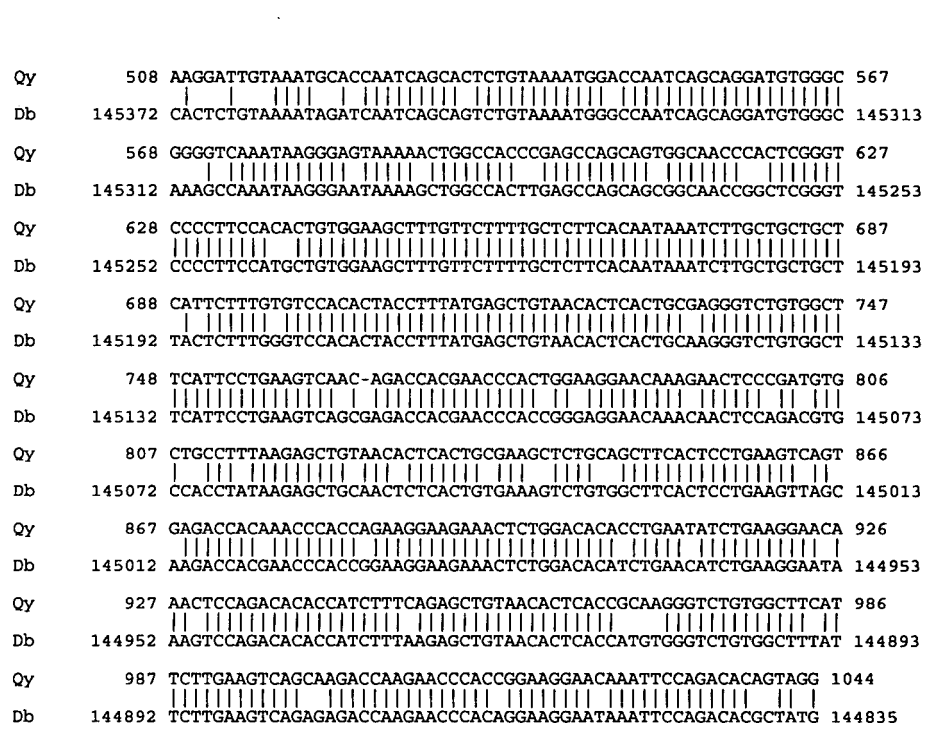
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.

FEATURES Location/Qualifiers
source 1. .169482
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP13-179C6"
/clone_lib="RPCI-13.1"

ORIGIN

Query Match 38.9%; Score 778; DB 8; Length 169482;
Best Local Similarity 85.7%; Pred. No. 4.8e-196;
Matches 924; Conservative 0; Mismatches 120; Indels 34; Gaps 4;





RESULT 19
HS785G19
LOCUS HS785G19 115231 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP4-785G19 on chromosome
20q11.1-11.23 Contains the XPOTP1 gene for exportin tRNA (nuclear
export receptor for tRNAs) pseudogene 1, a UTRN (utrophin
(homologous to dystrophin)) pseudogene, the ASIP gene for agouti
(mouse)-signaling protein, the 3' end of AHCY gene for
S-adenosylhomocysteine hydrolase and three CpG islands, complete
sequence.
ACCESSION AL035458
VERSION AL035458.35 GI:6624641
KEYWORDS HTG; AHCY; ASIP; CpG island; exportin; UTRN; utrophin; XPOTP1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 115231)
AUTHORS Babbage,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
COMMENT On Dec 21, 1999 this sequence version replaced gi:6562427.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP4-785G19 is from the library RPCI-4 constructed by the group of

```
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .115231
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704G19785"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="q11.1-11.23"
                     /clone="RP4-785G19"
                     /clone_lib="RPCI-4"
     misc_feature    100
                     /note="Clone_right_end: RP1-64K7"
     gene            complement(58003. .60400)
                     /locus_tag="RP4-785G19.3-001"
                     /pseudo
     CDS             complement(58003. .60400)
                     /locus_tag="RP4-785G19.3-001"
                     /note="match: proteins: O43592"
                     /pseudo
                     /codon_start=1
     misc_feature    74336. .74365
                     /note="Tandem repeat. Single clone region. Assembly
                     confirmed by restriction digest data."
     misc_feature    74892. .74893
                     /note="Tandem repeat"
     gene            join(complement(AL356299.16:10398. .10517),
                     complement(94506. .94646),complement(79351. .79538))
                     /gene="AHCY"
                     /locus_tag="RP4-785G19.1-003"
     mRNA            join(complement(AL356299.16:10398. .10517),
                     complement(94506. .94646),complement(79351. .79538))
                     /gene="AHCY"
                     /locus_tag="RP4-785G19.1-003"
                     /note="match: ESTs: AI241007.1"
     polyA_site      complement(79351)
                     /gene="AHCY"
                     /locus_tag="RP4-785G19.1-003"
     gene            complement(90461. .91045)
                     /locus_tag="RP4-785G19.2-001"
                     /pseudo
     CDS             complement(90461. .91045)
                     /locus_tag="RP4-785G19.2-001"
                     /note="match: proteins: AAH07860 O43583 Q9CQJ6"
                     /pseudo
                     /codon_start=1
     gene            104873. .113852
                     /gene="ASIP"
                     /locus_tag="RP4-785G19.4-001"
     mRNA            join(104873. .105042,107340. .107401,113499. .113852)
                     /gene="ASIP"
                     /locus_tag="RP4-785G19.4-001"
                     /note="match: ESTs: AI220203 AI378919 AI698587.1
                     AW206665.1 W02039 W32238
                     match: cDNAs: L06451 L06941 X99692 Y09877"
     CDS             join(104883. .105042,107340. .107401,113499. .113675)
                     /gene="ASIP"
                     /locus_tag="RP4-785G19.4-001"
                     /standard_name="OTTHUMP00000030681"
                     /note="match: proteins: P42127"
                     /codon_start=1
                     /protein_id="CAB96679.1"
                     /db_xref="GI:8953446"
                     /db_xref="Genew:745"
                     /db_xref="GOA:P42127"
                     /db_xref="HSSP:1HYK"
                     /db_xref="InterPro:IPR007733"
                     /db_xref="UniProt/Swiss-Prot:P42127"
                     /translation="MDVTRLLLATLLVFLCFFTANSHLPPEEKLRDDRSLRSNSSVNL
                     LDVPSVSIVALNKKSKQIGRKAAEKKRSSKKEASMKKVVRPRTPLSAPCVATRNSCKP
                     PAPACCDPCASCQCRFFRSACSCRVLSLNC"
     misc_feature    115231
                     /note="Clone_right_end: RP4-785G19"
ORIGIN

  Query Match             38.7%;  Score 773.4;  DB 8;  Length 115231;
  Best Local Similarity   84.0%;  Pred. No. 8.1e-195;
  Matches  952;  Conservative    0;  Mismatches  156;  Indels   25;  Gaps    6;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||| |||||||||||||||||| ||||||||||||| ||||||||||||||||  ||
Db       1640 AAAGGTTTCTGAAATCAGACAATGGCTTTCAAACTCTTACACCAACCTCTGGAGTTCAGC 1699

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
               || |||||||||||||||||||||||||||| |||||||||||||| || | || ||  |
Db       1700 AACGTGGCTTCTCCCCTTTCTAGGTCCTGTGGCAGCCATCTTGCTATTACTTGCCTTCAG 1759

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db       1760 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 1819

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| || ||| |||||||||||||||||||||| | ||||||
Db       1820 ACAGATGGTCTTACAAATGGAATCCCGAATGAGCTCAACTAACAACTTCTACCAAGGACC 1879

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||| | | || |||||||||| |||||||||||| ||||||||||||||||||||
Db       1880 CCTGGACCAATCTGCGGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 1939

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              ||||||||||||||||||||||||||||||||||||||||||||| || |||||| | ||
Db       1940 AACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATTGCCC 1999

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||| |||||||||| ||||||| ||||||||||||||||||||
Db       2000 AATTCCCAACAGCAGTTGGGTTGTCCTGTTTAGAGGGGGTATTGAGAGGTGAAGCCAGCT 2059

Qy        420 GGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              || ||| ||    ||| ||| |||||||||||||||| |||||||||||| |||||||
Db       2060 GGACTTCCTCCATCAAGTGAGGACTTGGAGAACTTTTCTGTCTAGCTAAAGGTTTGTAAA 2119

Qy        479 TGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTG 538
              |||||| ||||||||||||                ||||| |||||| ||| |||||
Db       2120 CGCACCAGTCAGCACTCTGT----------------AAAATGAACCAATGAGCGCTCTG 2162

Qy        539 TAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCAC 598
              |||||||||||||||||||||  ||||| ||| |||||||||| ||||| ||||||||
Db       2163 TAAAATGGACCAATCAGCAGGACATGGGCAGGGCCAAATAAGGGAATAAAAGCTGGCCAC 2222

Qy        599 CCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTT 658
              | |||||| |||||||||||  |||||||||||||||||| ||||||||||||||||||||
Db       2223 CTGAGCCATCAGTGGCAACCCGTTCGGGTCCCCTTCCACAGTGTGGAAGCTTTGTTCTTT 2282

Qy        659 TGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGC 718
                |||||||||||||||||||||||||||||| |||||| ||    ||||||||||||||||
Db       2283 CACTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTTAGCACTACCTTTATGAGC 2342

Qy        719 TGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCC 777
              |||||||||||  |||||||||||| | |||||| ||||||| | ||||||| |||||
Db       2343 TGTAACACTCACCACGAGGGTCTGTGGTTCCATTCCAGAAGTCAGCAAGACCACAAACCC 2402
```

```
Qy        778 ACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGA 837
              ||||| | ||||||| |||||| || ||||  ||||||| ||||||||||||| |||||||
Db       2403 ACTGGGAAGAACAAACAACTCCAGACGTGCCACCTTTAACAGCTGTAACACTGACTGCGA 2462

Qy        838 AGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACT 897
              || ||||  |||||||| |||||||||  ||||||| ||||||||||||||||||||||
Db       2463 AGGTCTGTGGCTTCACTTCTGAAGTCAGCAAGACCACGAACCCACCAGAAGGAAGAAACT 2522

Qy        898 CTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTA 957
              ||||||||| ||||| ||||    |||||||||||||||||| ||||||| ||| |||||
Db       2523 CTGGACACATCTGAACATCT----GAACAAACTCCAGACACAACATCTTTAAGAACTGTA 2578

Qy        958 ACACTCACCGCAAGGGTCTGTGGCTTCATTCTTG-AAGTCAGCAAGACCAAGAACCCACC 1016
              ||||||||| | |||||| |  |||||||||||| |||||||| ||||||||||||||||
Db       2579 ACACTCACCACGAGGGTCCGCAGCTTCATTCTTGAAAGTCAGCGAGACCAAGAACCCACC 2638

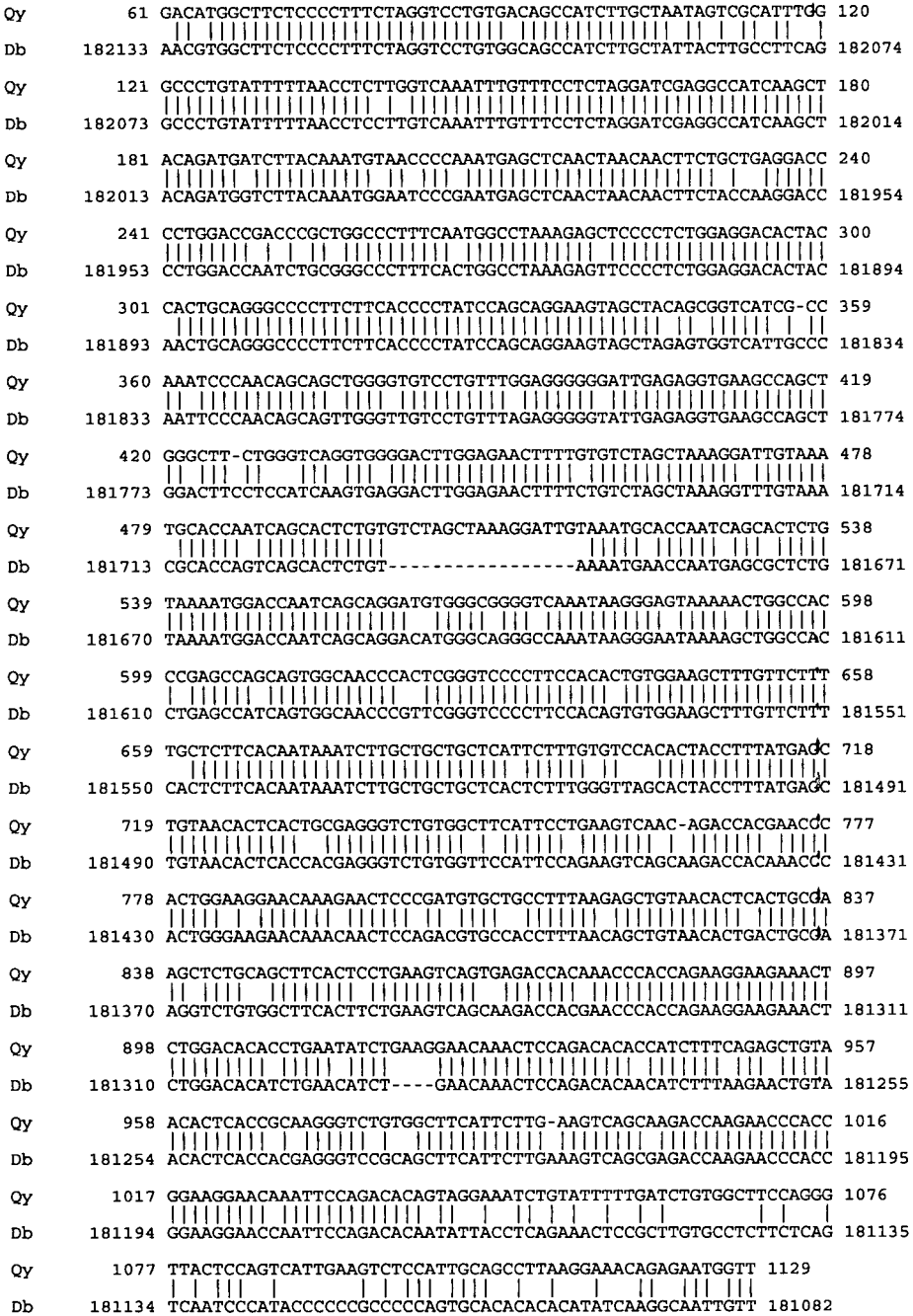
Qy       1017 GGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGG 1076
              ||||||||| |||||||||||| ||  |  || | |  | | |         | |   |
Db       2639 GGAAGGAACCAATTCCAGACACAATATTACCTCAGAAACTCCGCTTGTGCCTCTTCTCAG 2698

Qy       1077 TTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTT 1129
              | | |||  |       | | ||| ||||  |  |   |   ||  ||| |||
Db       2699 TCAATCCCATACCCCCCGCCCCCAGTGCACACACACATATCAAGGCAATTGTT 2751


RESULT 20
AL137253/c
LOCUS       AL137253              182753 bp    DNA     linear   HTG 13-JUN-2001
DEFINITION  Homo sapiens chromosome 20 clone RP11-263P6, 11 unordered pieces.
ACCESSION   AL137253
VERSION     AL137253.8  GI:10039490
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 8, 2000 this sequence version replaced gi:9931246.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA263P6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 179355 bases at least Q40
            Consensus quality: 180339 bases at least Q30
            Consensus quality: 180968 bases at least Q20
            Insert size: 181753; sum-of-contigs
            Insert size: 183363; 9.1% error; agarose-fp
            Quality coverage: 6.07x in Q20 bases; sum-of-contigs Quality
            coverage: 6.10x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11896: contig of 11896 bp in length
            *    11897    11996: gap of 100 bp
            *    11997    17768: contig of 5772 bp in length
            *    17769    17868: gap of 100 bp
            *    17869    33283: contig of 15415 bp in length
            *    33284    33383: gap of 100 bp
            *    33384    37004: contig of 3621 bp in length
            *    37005    37104: gap of 100 bp
            *    37105    72676: contig of 35572 bp in length
            *    72677    72776: gap of 100 bp
            *    72777    97002: contig of 24226 bp in length
            *    97003    97102: gap of 100 bp
            *    97103   121271: contig of 24169 bp in length
            *   121272   121371: gap of 100 bp
            *   121372   132660: contig of 11289 bp in length
            *   132661   132760: gap of 100 bp
            *   132761   145195: contig of 12435 bp in length
            *   145196   145295: gap of 100 bp
            *   145296   174784: contig of 29489 bp in length
            *   174785   174884: gap of 100 bp
            *   174885   182753: contig of 7869 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182753
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP11-263P6"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .11896
                     /note="assembly_fragment:02743
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    11997. .17768
                     /note="assembly_fragment:00211
                     fragment_chain:1"
     misc_feature    17869. .33283
                     /note="assembly_fragment:00124
                     fragment_chain:1"
     misc_feature    33384. .37004
                     /note="assembly_fragment:02657
                     fragment_chain:1"
     misc_feature    37105. .72676
                     /note="assembly_fragment:01648
                     fragment_chain:2"
     misc_feature    72777. .97002
                     /note="assembly_fragment:02959
                     fragment_chain:2"
     misc_feature    97103. .121271
                     /note="assembly_fragment:00003
                     fragment_chain:2"
     misc_feature    121372. .132660
                     /note="assembly_fragment:02923
                     fragment_chain:2"
     misc_feature    132761. .145195
                     /note="assembly_fragment:02105
                     fragment_chain:3"
     misc_feature    145296. .174784
                     /note="assembly_fragment:02149
                     fragment_chain:3"
     misc_feature    174885. .182753
                     /note="assembly_fragment:02331
                     fragment_chain:3
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match           38.7%;  Score 773.4;  DB 14;  Length 182753;
  Best Local Similarity   84.0%;  Pred. No. 8e-195;
  Matches  952;  Conservative    0;  Mismatches  156;  Indels   25;  Gaps    6;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||| |||||||||||||||| |||||||||||| ||||||||||||||||  ||
Db     182193 AAAGGTTTCTGAAATCAGACAATGGCTTTCAAACTCTTACACCAACCTCTGGAGTTCAGC 182134
```



```
RESULT 21
AC084395
LOCUS       AC084395              174084 bp    DNA     linear   HTG 30-OCT-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-260L15, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC084395
VERSION     AC084395.1  GI:11039612
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 174084)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174084)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0260L15
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 166512 bases at least Q40
            Consensus quality: 169389 bases at least Q30
            Consensus quality: 170542 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 172284; sum-of-contigs
            Quality coverage: 4.51 in Q20 bases; agarose-fp
            Quality coverage: 4.79 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      7746: contig of 7746 bp in length
            *     7747      7846: gap of unknown length
            *     7847     16751: contig of 8905 bp in length
            *    16752     16851: gap of unknown length
            *    16852     23977: contig of 7126 bp in length
            *    23978     24077: gap of unknown length
            *    24078     34459: contig of 10382 bp in length
            *    34460     34559: gap of unknown length
            *    34560     44597: contig of 10038 bp in length
            *    44598     44697: gap of unknown length
            *    44698     55249: contig of 10552 bp in length
            *    55250     55349: gap of unknown length
            *    55350     66281: contig of 10932 bp in length
            *    66282     66381: gap of unknown length
            *    66382     80808: contig of 14427 bp in length
            *    80809     80908: gap of unknown length
            *    80909     91732: contig of 10824 bp in length
            *    91733     91832: gap of unknown length
            *    91833    106328: contig of 14496 bp in length
            *   106329    106428: gap of unknown length
```

```
           *   106429   122673: contig of 16245 bp in length
           *   122674   122773: gap of unknown length
           *   122774   140593: contig of 17820 bp in length
           *   140594   140693: gap of unknown length
           *   140694   143092: contig of 2399 bp in length
           *   143093   143192: gap of unknown length
           *   143193   144628: contig of 1436 bp in length
           *   144629   144728: gap of unknown length
           *   144729   148312: contig of 3584 bp in length
           *   148313   148412: gap of unknown length
           *   148413   153703: contig of 5291 bp in length
           *   153704   153803: gap of unknown length
           *   153804   158969: contig of 5166 bp in length
           *   158970   159069: gap of unknown length
           *   159070   164878: contig of 5809 bp in length
           *   164879   164978: gap of unknown length
           *   164979   174084: contig of 9106 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174084
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-260L15"
     misc_feature    1. .7746
                     /note="assembly_name:Contig10"
     gap             7747. .7846
                     /estimated_length=unknown
     misc_feature    7847. .16751
                     /note="assembly_name:Contig11"
     gap             16752. .16851
                     /estimated_length=unknown
     misc_feature    16852. .23977
                     /note="assembly_name:Contig12"
     gap             23978. .24077
                     /estimated_length=unknown
     misc_feature    24078. .34459
                     /note="assembly_name:Contig13"
     gap             34460. .34559
                     /estimated_length=unknown
     misc_feature    34560. .44597
                     /note="assembly_name:Contig14"
     gap             44598. .44697
                     /estimated_length=unknown
     misc_feature    44698. .55249
                     /note="assembly_name:Contig15"
     gap             55250. .55349
                     /estimated_length=unknown
     misc_feature    55350. .66281
                     /note="assembly_name:Contig16"
     gap             66282. .66381
                     /estimated_length=unknown
     misc_feature    66382. .80808
                     /note="assembly_name:Contig17"
     gap             80809. .80908
                     /estimated_length=unknown
     misc_feature    80909. .91732
                     /note="assembly_name:Contig18"
     gap             91733. .91832
                     /estimated_length=unknown
     misc_feature    91833. .106328
                     /note="assembly_name:Contig19"
     gap             106329. .106428
                     /estimated_length=unknown
     misc_feature    106429. .122673
                     /note="assembly_name:Contig20"
     gap             122674. .122773
                     /estimated_length=unknown
     misc_feature    122774. .140593
                     /note="assembly_name:Contig21"
     gap             140594. .140693
                     /estimated_length=unknown
     misc_feature    140694. .143092
                     /note="assembly_name:Contig3"
     gap             143093. .143192
                     /estimated_length=unknown
     misc_feature    143193. .144628
                     /note="assembly_name:Contig4"
     gap             144629. .144728
                     /estimated_length=unknown
     misc_feature    144729. .148312
                     /note="assembly_name:Contig5"
     gap             148313. .148412
                     /estimated_length=unknown
     misc_feature    148413. .153703
                     /note="assembly_name:Contig6
                     clone_end:SP6
                     vector_side:right"
     gap             153704. .153803
                     /estimated_length=unknown
     misc_feature    153804. .158969
                     /note="assembly_name:Contig7
                     clone_end:T7
                     vector_side:left"
     gap             158970. .159069
                     /estimated_length=unknown
     misc_feature    159070. .164878
                     /note="assembly_name:Contig8"
     gap             164879. .164978
                     /estimated_length=unknown
     misc_feature    164979. .174084
                     /note="assembly_name:Contig9"
ORIGIN

  Query Match             38.6%;  Score 772.6;  DB 14;  Length 174084;
  Best Local Similarity   86.4%;  Pred. No. 1.3e-194;
  Matches  922;  Conservative    0;  Mismatches  114;  Indels   31;  Gaps    5;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||| |||||||||||   ||||||| |||||||||| |||||||||| |||
Db      29454 AAAGGCTTCCAAAATCAGACAACATCTTTCAAGCTCTTATACCAGCCTCTGGAGTT-GGC 29512

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||||  || ||||||||||||| || || | || ||
Db      29513 AACATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTATTACTCACCTTCGG 29572

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||| |||||||||||||| | |||||||||||||||||||||||||||||| ||||||||
Db      29573 GCCCGGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATCGAGGCTATCAAGCT 29632

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              |||||||||||||||||| ||||||||||||||||||||||||   || |  ||||||
Db      29633 ACAGATGATCTTACAAATGGAACCCCAAATGAGCTCAACTAACAA---CTACCAAGGACC 29689

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||| |||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      29690 CCTGGACCAACCCGCTGGCCCTTTCAATGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 29749

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              ||||||||| || |||||  |||||||||||||||||||||||| || |||||||  ||
Db      29750 AACTGCAGGGTCCTTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATCACCC 29809

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||  |||| | |||| ||||||||||||||||||||
Db      29810 AATTCCCAACAGCAGTTGGGGTGTCTTGTTAAGTGGGGAGATTGAGAGGTGAAGCCAGCT 29869

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||  |||||||||||||||||||||| |||||||||| ||||||||||
Db      29870 GGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAGAGGATTGTAAAC 29929

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGC---------------- 522
              ||||||||||  |||||||||||||| ||| ||||||||||
Db      29930 ACACCAATCAGTGCTCTGTGTCTAGCTAGAGGTTTGTAAATGCACCAATCAGCACTCTGT 29989

Qy        523 --------ACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCA 574
                       |||||||||||||||||||||||||||||||||| |||||| |||| ||| ||
```

```
Db      29990 AAAAACGGACCAATCAGCACTCTGTAAAATGGACCAATCAGTAGGATGCGGGCAGGGCCA 30049

Qy        575 AATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTC 634
              ||||||||| ||||| ||||||||| ||| ||||||||||||||||||||||||||||||
Db      30050 AATAAGGGAATAAAAGCTGGCCACCTGAGTCAGCAGTGGCAACCCACTCGGGTCCCCTTC 30109

Qy        635 CACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTT 694
              ||  |||||||||||||||||||||  |||||||||||||||||||||||||||||| ||||
Db      30110 CATGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCTGCTCACTCTT 30169

Qy        695 TGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCC 754
              || |||||||| |||||||||||||| |||||||||||||| |||||    |||||| |||
Db      30170 TGGGTCCACACCACCTTTATGAGCTGCAACACTCACTGCGAAGGTCTTCAGCTTCACTCC 30229

Qy        755 TGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813
              |||||||| | |||||||||||||||  || ||||||||   ||| | ||  |||  |||| 
Db      30230 TGAAGTCAGCGAGACCACGAACCCATGGGGAGGAACAATCGACTTCAGACATGCCACCTT 30289

Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873
              |||||||||||||||||||||||||||| ||||  |||||||||||||||||||  ||||||
Db      30290 TAAGAGCTGTAACACTCACTGCGAAGGTCTGTGGCTTCACTCCTGAAGTCAGCAAGACCA 30349

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933
              | |||||||  |||||||||| |  ||||||| ||||| ||||||| ||||||||||
Db      30350 CGAACCCACTGGAAGGAAGAAATTTCGGACACATCTGAACATCTGAATGAACAAACTCTG 30409

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
              ||||| |||||||| ||| |||||||||||| |  |||||   ||| ||||||||||||
Db      30410 GACACGCCATCTTTAAGAACTGTAACACTCACTGTGAGGGTTCATGGGTTCATTCTTGAA 30469

Qy        994 GTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAG 1040
              |||||||||||||||||||||||| |||||||| ||||||| |||||||
Db      30470 GTCAGCAAGACCAAGAACCCACCAGAAGGAACCAATTCCGGACACAG 30516
```

RESULT 22
AL160399
LOCUS AL160399 126595 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-161N21 on chromosome 6, complete sequence.
ACCESSION AL160399
VERSION AL160399.13 GI:14018252
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 126595)
AUTHORS Holt,K.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On May 11, 2001 this sequence version replaced gi:13751310.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6
RP11-161N21 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vega@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
FEATURES Location/Qualifiers
source 1. .126595
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-161N21"
/clone_lib="RPCI-11.1"
misc_feature 100
/note="Clone_right_end: RP3-343I24"
misc_feature 17528. .17592
/note="Sequence from clone PCR only. Sequence from overlapping clone bA76I19 (AL355854). Assembly confirmed by restriction digest data. Sequence from uni-directional dGTP big dye terminator reads only."
misc_feature 17753. .17777
/note="Sequence from clone PCR only. Sequence from overlapping clone bA76I19 (AL355854). Assembly confirmed by restriction digest data. Sequence from uni-directional dGTP big dye terminator reads only."
misc_feature 126496
/note="Clone_left_end: RP3-493I16"
ORIGIN

Query Match 38.6%; Score 771.8; DB 8; Length 126595;
Best Local Similarity 86.2%; Pred. No. 2.2e-194;
Matches 899; Conservative 0; Mismatches 117; Indels 27; Gaps 3;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56831 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 56890

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
               || |||||||||||||||||||||||    | |||  ||||||||| || | || |||||
Db      56891 AACGTGGCTTCTCCCCTTTCTAGGTCCCACGGCAGTTATCTTGCTATTACTTGCCTTTGG 56950

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              || || |||||||||||| | ||||||||||||||| ||| || ||||||||||||||||
Db      56951 GCACTATATTTTTAACCTCCTTGTCAAATTTGTTTCCTGTAGAATTGAGGCCATCAAGCT 57010

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||||  |||| |||||||||||||||||||||| | |||||||
Db      57011 ACAGATGGTCTTACAAATGGTACCCTAAATGAGCTCAACTAACAACTTCTACCGAGGACC 57070

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||| |||| ||| |||||||| |||||||||||  ||||||||||||||||||||
Db      57071 CCTGGACCAACCCACTGTCCCTTTCACTGGCCTAAAGAATTCCCCTCTGGAGGACACTAC 57130

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||||||||||||| |||||||| |||||||||||||| || ||||||| |||
Db      57131 AACTGCAGGGCCCCTTCTTCGCCCCTATCTAGCAGGAAGTAGCTAGAGTGGTCATCGGCC 57190

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||  |||||||||||||||  ||||||  ||||||||||||||  |||
Db      57191 AATTCCCAACAGCATTTGGGGTGTCCTGTTTAGAGGGGTTATTGAGAGGTGAAGCTGGCT 57250

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||| |||||||||| |||||||||| ||||||||||||||||||||||||
Db      57251 GGGCTTCTGGGTCGGGTGGGGACTTAGAGAACTTTTCTGTCTAGCTAAAGGATTGTAAAC 57310

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGC---------------- 522
              |||||||||  |||||||||||||||||||| ||||||| ||
Db      57311 ACACCAATCAGTGCTCTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATAAGCACTCTGT 57370

Qy        523 --------ACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCA 574
```

```
                  |||||| |||||||||||||||||||||||||||||||||||||||||| |||| ||
Db      57371 AAAAATAGACCAATAAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCA 57430

Qy        575 AATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTC 634
              ||||||||| ||||| |||||||||| |||||||||  |||||||  || |||||| ||||
Db      57431 AATAAGGGAATAAAAGCTGGCCACCTGAGCCAGCAGCAGCAACCCGCTGGGGTCCTCTTC 57490

Qy        635 CACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTT 694
              ||| ||||| |||||||||||||||||||||  ||||||||||||||||||||| ||||
Db      57491 CACGCTGTGAAAGCTTTGTTCTTTTGCTCTTTGCAATAAATCTTGCTGCTGCTCACTCTT 57550

Qy        695 TGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCC 754
              || | |  |||||||||||||||||||||||||||||| | ||||||  |||||  |||
Db      57551 TGGGCCTGCACTACCTTTATGAGCTGTAACACTCACTGCAAAGGTCTGCAGCTTCTCTCC 57610

Qy        755 TGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813
              ||||| || | ||||||| |||||||||| || ||  || |||||  ||  |||  ||||
Db      57611 TGAAGCCAGCGAGACCACAAACCCACTGGGAGAAATGAACAACTCTGGACATGCCACCTT 57670

Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873
              |||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||
Db      57671 TAAGAGCTGTAACACTCACTGTGAAGGTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 57730

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933
                |||||||||||||||||||||| ||||||   |||| ||||||||||||||| ||
Db      57731 GGAACCCACCAGAAGGAAGAAACTCGGGACACCTGTGAACATCTGAAGGAACAAACGCCG 57790

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
              | ||||| |||||| | | |||||| ||||||||  ||||||  ||||||||||||||||
Db      57791 GGCACACTATCTTTAACAACTGTAAGACTCACCGTGAGGGTCCATGGCTTCATTCTTGAA 57850

Qy        994 GTCAGCAAGACCAAGAACCCACC 1016
              |||| | |||| ||||||||||
Db      57851 GTCACCGAGACTAAGAACCCACC 57873


RESULT 23
AL355854
LOCUS       AL355854              164532 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-76I19, 11 unordered pieces.
ACCESSION   AL355854
VERSION     AL355854.8  GI:10186514
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2000 this sequence version replaced gi:9931763.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA76I19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 160164 bases at least Q40
            Consensus quality: 161830 bases at least Q30
            Consensus quality: 162686 bases at least Q20
            Insert size: 163532; sum-of-contigs
            Insert size: 167031; 6.6% error; agarose-fp
            Quality coverage: 4.89x in Q20 bases; sum-of-contigs Quality
            coverage: 5.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5323: contig of 5323 bp in length
            *     5324     5423: gap of 100 bp
            *     5424    18444: contig of 13021 bp in length
            *    18445    18544: gap of 100 bp
            *    18545    22423: contig of 3879 bp in length
            *    22424    22523: gap of 100 bp
            *    22524    54775: contig of 32252 bp in length
            *    54776    54875: gap of 100 bp
            *    54876    79765: contig of 24890 bp in length
            *    79766    79865: gap of 100 bp
            *    79866   101811: contig of 21946 bp in length
            *   101812   101911: gap of 100 bp
            *   101912   120185: contig of 18274 bp in length
            *   120186   120285: gap of 100 bp
            *   120286   129687: contig of 9402 bp in length
            *   129688   129787: gap of 100 bp
            *   129788   132901: contig of 3114 bp in length
            *   132902   133001: gap of 100 bp
            *   133002   154594: contig of 21593 bp in length
            *   154595   154694: gap of 100 bp
            *   154695   164532: contig of 9838 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164532
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-76I19"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .5323
                     /note="assembly_fragment:01547
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    5424. .18444
                     /note="assembly_fragment:00390
                     fragment_chain:1"
     misc_feature    18545. .22423
                     /note="assembly_fragment:00760
                     fragment_chain:1"
     misc_feature    22524. .54775
                     /note="assembly_fragment:00977
                     fragment_chain:1"
     misc_feature    54876. .79765
                     /note="assembly_fragment:01205
                     fragment_chain:1"
     misc_feature    79866. .101811
                     /note="assembly_fragment:02190
                     fragment_chain:2"
     misc_feature    101912. .120185
                     /note="assembly_fragment:01164
                     fragment_chain:2"
     misc_feature    120286. .129687
                     /note="assembly_fragment:00323"
     misc_feature    129788. .132901
                     /note="assembly_fragment:00850"
     misc_feature    133002. .154594
                     /note="assembly_fragment:01231
                     fragment_chain:3"
     misc_feature    154695. .164532
                     /note="assembly_fragment:01221
                     fragment_chain:3
                     clone_end:SP6
                     vector_side:right"
```

ORIGIN

Query Match 38.6%; Score 771.8; DB 14; Length 164532;
Best Local Similarity 86.2%; Pred. No. 2.2e-194;
Matches 899; Conservative 0; Mismatches 117; Indels 27; Gaps 3;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
Db     115346 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 115405

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
Db     115406 AACGTGGCTTCTCCCCTTTCTAGGTCCCACGGCAGTTATCTTGCTATTACTTGCCTTTGG 115465

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
Db     115466 GCACTATATTTTTAACCTCCTTGTCAAATTTGTTTCCTGTAGAATTGAGGCCATCAAGCT 115525

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
Db     115526 ACAGATGGTCTTACAAATGGTACCCTAAATGAGCTCAACTAACAACTTCTACCGAGGACC 115585

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
Db     115586 CCTGGACCAACCCACTGTCCCTTTCACTGGCCTAAAGAATTCCCCTCTGGAGGACACTAC 115645

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
Db     115646 AACTGCAGGGCCCCTTCTTCGCCCCTATCTAGCAGGAAGTAGCTAGAGTGGTCATCGGCC 115705

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
Db     115706 AATTCCCAACAGCATTTGGGGTGTCCTGTTTAGAGGGGTTATTGAGAGGTGAAGCTGGCT 115765

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
Db     115766 GGGCTTCTGGGTCGGGTGGGGACTTAGAGAACTTTTCTGTCTAGCTAAAGGATTGTAAAC 115825

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGC---------------- 522
Db     115826 ACACCAATCAGTGCTCTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATAAGCACTCTGT 115885

Qy        523 --------ACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCA 574
Db     115886 AAAAATAGACCAATAAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCA 115945

Qy        575 AATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTC 634
Db     115946 AATAAGGGAATAAAAGCTGGCCACCTGAGCCAGCAGCAGCAACCCGCTGGGGTCCTCTTC 116005

Qy        635 CACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTT 694
Db     116006 CACGCTGTGAAAGCTTTGTTCTTTTGCTCTTTGCAATAAATCTTGCTGCTGCTCACTCTT 116065

Qy        695 TGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCC 754
Db     116066 TGGGCCTGCACTACCTTTATGAGCTGTAACACTCACTGCAAAGGTCTGCAGCTTCTCTCC 116125

Qy        755 TGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813
Db     116126 TGAAGCCAGCGAGACCACAAACCCACTGGGAGAAATGAACAACTCTGGACATGCCACCTT 116185

Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873
Db     116186 TAAGAGCTGTAACACTCACTGTGAAGGTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 116245

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933
Db     116246 GGAACCCACCAGAAGGAAGAAACTCGGGACACCTGTGAACATCTGAAGGAACAAACGCCG 116305

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
Db     116306 GGCACACTATCTTTAACAACTGTAAGACTCACCGTGAGGGTCCATGGCTTCATTCTTGAA 116365

Qy        994 GTCAGCAAGACCAAGAACCCACC 1016
Db     116366 GTCACCGAGACTAAGAACCCACC 116388
```

RESULT 24
AP000924
LOCUS AP000924 172274 bp DNA linear PRI 17-JUN-2003
DEFINITION Homo sapiens genomic DNA, chromosome 11 clone:RP11-758F15, complete sequence.
ACCESSION AP000924
VERSION AP000924.7 GI:31790759
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (1999)
REFERENCE 2 (bases 1 to 172274)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Jun 16, 2003 this sequence version replaced gi:13359382.
FEATURES Location/Qualifiers
source 1. .172274
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-758F15"

ORIGIN

Query Match 38.5%; Score 771; DB 8; Length 172274;
Best Local Similarity 86.3%; Pred. No. 3.5e-194;
Matches 921; Conservative 0; Mismatches 115; Indels 31; Gaps 5;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
Db      39185 AAAGGCTTCCAAAATCAGACAACATCTTTCAAGCTCTTATACCAGCCTCTGGAGTT-GGC 39243

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
Db      39244 AACATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTATTACTCACCTTCGG 39303

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
Db      39304 GCCCGGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATCGAGGCTATCAAGCT 39363

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
Db      39364 ACAGATGATCTTACAAATGGAACCCCAAATGAGCTCAACTAACAA---CTACCAAGGACC 39420

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
Db      39421 CCTGGACCAACCCGCTGGCCCTTTCAATGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 39480

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
Db      39481 AACTGCAGGGTCCTTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATCACCC 39540

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
```

```
Db      39541 AATTCCCAACAGCAGTTGGGGTGTCTTGTTAAGTGGGGAGATTGAGAGGTGAAGCCAGCT 39600

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479

Db      39601 GGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAGAGGATTGTAAAC 39660

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGC---------------- 522

Db      39661 ACACCAATCAGTGCTATGTGTCTAGCTAGAGGTTTGTAAATGCACCAATCAGCACTCTGT 39720

Qy        523 --------ACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCA 574

Db      39721 AAAAACGGACCAATCAGCACTCTGTAAAATGGACCAATCAGTAGGATGCGGGCAGGGCCA 39780

Qy        575 AATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTC 634

Db      39781 AATAAGGGAATAAAAGCTGGCCACCTGAGTCAGCAGTGGCAACCCACTCGGGTCCCCTTC 39840

Qy        635 CACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTT 694

Db      39841 CATGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCTGCTCACTCTT 39900

Qy        695 TGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCC 754

Db      39901 TGGGTCCACACCACCTTTATGAGCTGCAACACTCACTGCGAAGGTCTTCAGCTTCACTCC 39960

Qy        755 TGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813

Db      39961 TGAAGTCAGCGAGACCACGAACCCATGGGGAGGAACAATCGACTTCAGACATGCCACCTT 40020

Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873

Db      40021 TAAGAGCTGTAACACTCACTGCGAAGGTCTGTGGCTTCACTCCTGAAGTCAGCAAGACCA 40080

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933

Db      40081 CGAACCCACTGGAAGGAAGAAATTTCGGACACATCTGAACATCTGAATGAACAAACTCTG 40140

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993

Db      40141 GACACGCCATCTTTAAGAACTGTAACACTCACTGTGAGGGTTCATGGGTTCATTCTTGAA 40200

Qy        994 GTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAG 1040

Db      40201 GTCAGCAAGACCAAGAACCCACCAGAAGGAACCAATTCCGGACACAG 40247
```

RESULT 25
AP002788
LOCUS AP002788 174311 bp DNA linear PRI 15-MAR-2003
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-347E10, complete sequence.
ACCESSION AP002788
VERSION AP002788.3 GI:13429927
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 174311)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Mar 21, 2001 this sequence version replaced gi:9845043.
FEATURES Location/Qualifiers
source 1. .174311
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-347E10"
ORIGIN

Query Match 38.5%; Score 771; DB 8; Length 174311;
Best Local Similarity 86.3%; Pred. No. 3.5e-194;
Matches 921; Conservative 0; Mismatches 115; Indels 31; Gaps 5;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60

Db     147400 AAAGGCTTCCAAAATCAGACAACATCTTTCAAGCTCTTATACCAGCCTCTGGAGTT-GGC 147458

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120

Db     147459 AACATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTATTACTCACCTTCGG 147518

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180

Db     147519 GCCCGGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATCGAGGCTATCAAGCT 147578

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240

Db     147579 ACAGATGATCTTACAAATGGAACCCCAAATGAGCTCAACTAACAA---CTACCAAGGACC 147635

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300

Db     147636 CCTGGACCAACCCGCTGGCCCTTTCAATGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 147695

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359

Db     147696 AACTGCAGGGTCCTTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATCACCC 147755

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419

Db     147756 AATTCCCAACAGCAGTTGGGGTGTCTTGTTAAGTGGGGAGATTGAGAGGTGAAGCCAGCT 147815

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479

Db     147816 GGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAGAGGATTGTAAAC 147875

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGC---------------- 522

Db     147876 ACACCAATCAGTGCTATGTGTCTAGCTAGAGGTTTGTAAATGCACCAATCAGCACTCTGT 147935

Qy        523 --------ACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCA 574

Db     147936 AAAAACGGACCAATCAGCACTCTGTAAAATGGACCAATCAGTAGGATGCGGGCAGGGCCA 147995

Qy        575 AATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTC 634

Db     147996 AATAAGGGAATAAAAGCTGGCCACCTGAGTCAGCAGTGGCAACCCACTCGGGTCCCCTTC 148055

Qy        635 CACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTT 694

Db     148056 CATGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCTGCTCACTCTT 148115

Qy        695 TGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCC 754

Db     148116 TGGGTCCACACCACCTTTATGAGCTGCAACACTCACTGCGAAGGTCTTCAGCTTCACTCC 148175

Qy        755 TGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813

Db     148176 TGAAGTCAGCGAGACCACGAACCCATGGGGAGGAACAATCGACTTCAGACATGCCACCTT 148235
```

```
Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873
              ||||||||||||||||||||||||| ||||  |||||||||||||||||||  ||||||
Db     148236 TAAGAGCTGTAACACTCACTGCGAAGGTCTGTGGCTTCACTCCTGAAGTCAGCAAGACCA 148295

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933
              | |||||||  |||||||||| |  ||||||| ||||| ||||||| ||||||||||
Db     148296 CGAACCCACTGGAAGGAAGAAATTTCGGACACATCTGAACATCTGAATGAACAAACTCTG 148355

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
              ||||| ||||||| ||| ||||||||||||| |  |||||   ||| ||||||||||||
Db     148356 GACACGCCATCTTTAAGAACTGTAACACTCACTGTGAGGGTTCATGGGTTCATTCTTGAA 148415

Qy        994 GTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAG 1040
              |||||||||||||||||||||| |||||||| ||||||| |||||||
Db     148416 GTCAGCAAGACCAAGAACCCACCAGAAGGAACCAATTCCGGACACAG 148462


RESULT 26
AC015806/c
LOCUS       AC015806               190592 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-420H22 map 11, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC015806
VERSION     AC015806.3  GI:8096897
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 190592)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-420H22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190592)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 26, 2000 this sequence version replaced gi:6692339.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2227
                Center clone name: 420_H_22
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 184180 bases at least Q40
                Consensus quality: 187135 bases at least Q30
                Consensus quality: 188337 bases at least Q20
                Insert size: 184000; agarose-fp
                Insert size: 189692; sum-of-contigs
                Quality coverage: 5.8 in Q20 bases; agarose-fp
                Quality coverage: 5.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3025: contig of 3025 bp in length
            *     3026     3125: gap of 100 bp
            *     3126     8902: contig of 5777 bp in length
            *     8903     9002: gap of 100 bp
            *     9003    16758: contig of 7756 bp in length
            *    16759    16858: gap of 100 bp
            *    16859    33120: contig of 16262 bp in length
            *    33121    33220: gap of 100 bp
            *    33221    49787: contig of 16567 bp in length
            *    49788    49887: gap of 100 bp
            *    49888    68295: contig of 18408 bp in length
            *    68296    68395: gap of 100 bp
            *    68396    93348: contig of 24953 bp in length
            *    93349    93448: gap of 100 bp
            *    93449   127332: contig of 33884 bp in length
            *   127333   127432: gap of 100 bp
            *   127433   156645: contig of 29213 bp in length
            *   156646   156745: gap of 100 bp
            *   156746   190592: contig of 33847 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190592
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-420H22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .3025
                     /note="assembly_fragment"
     gap             3026. .3125
                     /estimated_length=100
     misc_feature    3126. .8902
                     /note="assembly_fragment"
     gap             8903. .9002
                     /estimated_length=100
     misc_feature    9003. .16758
                     /note="assembly_fragment"
     gap             16759. .16858
                     /estimated_length=100
     misc_feature    16859. .33120
                     /note="assembly_fragment"
     gap             33121. .33220
                     /estimated_length=100
     misc_feature    33221. .49787
                     /note="assembly_fragment"
     gap             49788. .49887
                     /estimated_length=100
     misc_feature    49888. .68295
                     /note="assembly_fragment"
     gap             68296. .68395
                     /estimated_length=100
     misc_feature    68396. .93348
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     gap             93349. .93448
                     /estimated_length=100
     misc_feature    93449. .127332
                     /note="assembly_fragment"
     gap             127333. .127432
                     /estimated_length=100
```

```
     misc_feature    127433. .156645
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     gap             156646. .156745
                     /estimated_length=100
     misc_feature    156746. .190592
                     /note="assembly_fragment"
ORIGIN
```

Query Match 38.5%; Score 771; DB 14; Length 190592;
Best Local Similarity 86.3%; Pred. No. 3.5e-194;
Matches 921; Conservative 0; Mismatches 115; Indels 31; Gaps 5;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||| ||||||||||  ||||||| |||||||||| |||||||||| ||| |
Db      99230 AAAGGCTTCCAAAATCAGACAACATCTTTCAAGCTCTTATACCAGCCTCTGGAGTT-GGC 99172

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||  || ||||||||||||| || || | || ||
Db      99171 AACATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTATTACTCACCTTCGG 99112

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||| |||||||||||||| | |||||||||||||||||||||||||||||| ||||||||
Db      99111 GCCCGGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATCGAGGCTATCAAGCT 99052

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              |||||||||||||||||| |||||||||||||||||||||||   || |  ||||||
Db      99051 ACAGATGATCTTACAAATGGAACCCCAAATGAGCTCAACTAACAA---CTACCAAGGACC 98995

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      98994 CCTGGACCAACCCGCTGGCCCTTTCAATGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 98935

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||| || |||||  |||||||||||||||||||||||| || |||||||  ||
Db      98934 AACTGCAGGGTCCTTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATCACCC 98875

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||| ||||  | |||| ||||||||||||||||||||||
Db      98874 AATTCCCAACAGCAGTTGGGGTGTCTTGTTAAGTGGGGAGATTGAGAGGTGAAGCCAGCT 98815

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||  |||||||||||||||||||| |||||||||| ||||||||||
Db      98814 GGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAGAGGATTGTAAAC 98755

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGC---------------- 522
              ||||||||||  || |||||||||||| ||| ||||||||||
Db      98754 ACACCAATCAGTGCTATGTGTCTAGCTAGAGGTTTGTAAATGCACCAATCAGCACTCTGT 98695

Qy        523 --------ACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCA 574
                      |||||||||||||||||||||||||||||||||| |||||| |||| ||| ||
Db      98694 AAAAACGGACCAATCAGCACTCTGTAAAATGGACCAATCAGTAGGATGCGGGCAGGGCCA 98635

Qy        575 AATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTC 634
              |||||||| ||||| |||||||||| ||| ||||||||||||||||||||||||||||
Db      98634 AATAAGGGAATAAAAGCTGGCCACCTGAGTCAGCAGTGGCAACCCACTCGGGTCCCCTTC 98575

Qy        635 CACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTT 694
              ||  |||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db      98574 CATGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCTGCTCACTCTT 98515

Qy        695 TGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCC 754
              || |||||||| |||||||||||||| |||||||||||||| |||||   |||||| |||
Db      98514 TGGGTCCACACCACCTTTATGAGCTGCAACACTCACTGCGAAGGTCTTCAGCTTCACTCC 98455

Qy        755 TGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813
              |||||||| | |||||||||||||  || ||||||||   ||| | ||  |||  ||||
Db      98454 TGAAGTCAGCGAGACCACGAACCCATGGGGAGGAACAATCGACTTCAGACATGCCACCTT 98395

Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873
              |||||||||||||||||||||||||| ||||  |||||||||||||||||||  ||||||
Db      98394 TAAGAGCTGTAACACTCACTGCGAAGGTCTGTGGCTTCACTCCTGAAGTCAGCAAGACCA 98335

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933
              | |||||||  |||||||||| |  ||||||| ||||| ||||||| ||||||||||
Db      98334 CGAACCCACTGGAAGGAAGAAATTTCGGACACATCTGAACATCTGAATGAACAAACTCTG 98275

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
              ||||| |||||||| ||| |||||||||||||  |  |||||   ||| ||||||||||||
Db      98274 GACACGCCATCTTTAAGAACTGTAACACTCACTGTGAGGGTTCATGGGTTCATTCTTGAA 98215

Qy        994 GTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAG 1040
              |||||||||||||||||||||||| |||||||| |||||| |||||||
Db      98214 GTCAGCAAGACCAAGAACCCACCAGAAGGAACCAATTCCGGACACAG 98168
```

```
RESULT 27
AC024401
LOCUS       AC024401               168298 bp    DNA     linear   HTG 04-MAY-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-488M11 map 2, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC024401
VERSION     AC024401.2  GI:9994150
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 168298)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 2, clone RP11-488M11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168298)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 8, 2000 this sequence version replaced gi:7108197.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6777
                Center clone name: 488_M_11
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 163206 bases at least Q40
```

```
            Consensus quality: 166077 bases at least Q30
            Consensus quality: 166822 bases at least Q20
            Insert size: 168000; agarose-fp
            Insert size: 167098; sum-of-contigs
            Quality coverage: 5.3 in Q20 bases; agarose-fp
            Quality coverage: 5.3 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 13 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    19212: contig of 19212 bp in length
          *    19213    19312: gap of 100 bp
          *    19313    20607: contig of 1295 bp in length
          *    20608    20707: gap of 100 bp
          *    20708    24577: contig of 3870 bp in length
          *    24578    24677: gap of 100 bp
          *    24678    29497: contig of 4820 bp in length
          *    29498    29597: gap of 100 bp
          *    29598    33210: contig of 3613 bp in length
          *    33211    33310: gap of 100 bp
          *    33311    38616: contig of 5306 bp in length
          *    38617    38716: gap of 100 bp
          *    38717    47747: contig of 9031 bp in length
          *    47748    47847: gap of 100 bp
          *    47848    59600: contig of 11753 bp in length
          *    59601    59700: gap of 100 bp
          *    59701    72620: contig of 12920 bp in length
          *    72621    72720: gap of 100 bp
          *    72721    99437: contig of 26717 bp in length
          *    99438    99537: gap of 100 bp
          *    99538   126237: contig of 26700 bp in length
          *   126238   126337: gap of 100 bp
          *   126338   149961: contig of 23624 bp in length
          *   149962   150061: gap of 100 bp
          *   150062   168298: contig of 18237 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168298
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-488M11"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .19212
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             19213. .19312
                     /estimated_length=100
     misc_feature    19313. .20607
                     /note="assembly_fragment"
     gap             20608. .20707
                     /estimated_length=100
     misc_feature    20708. .24577
                     /note="assembly_fragment"
     gap             24578. .24677
                     /estimated_length=100
     misc_feature    24678. .29497
                     /note="assembly_fragment"
     gap             29498. .29597
                     /estimated_length=100
     misc_feature    29598. .33210
                     /note="assembly_fragment"
     gap             33211. .33310
                     /estimated_length=100
     misc_feature    33311. .38616
                     /note="assembly_fragment"
     gap             38617. .38716
                     /estimated_length=100
     misc_feature    38717. .47747
                     /note="assembly_fragment"
     gap             47748. .47847
                     /estimated_length=100
     misc_feature    47848. .59600
                     /note="assembly_fragment"
     gap             59601. .59700
                     /estimated_length=100
     misc_feature    59701. .72620
                     /note="assembly_fragment"
     gap             72621. .72720
                     /estimated_length=100
     misc_feature    72721. .99437
                     /note="assembly_fragment"
     gap             99438. .99537
                     /estimated_length=100
     misc_feature    99538. .126237
                     /note="assembly_fragment"
     gap             126238. .126337
                     /estimated_length=100
     misc_feature    126338. .149961
                     /note="assembly_fragment"
     gap             149962. .150061
                     /estimated_length=100
     misc_feature    150062. .168298
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match            38.0%;  Score 761.2;  DB 14;  Length 168298;
  Best Local Similarity  88.0%;  Pred. No. 1.4e-191;
  Matches  928;  Conservative    0;  Mismatches   98;  Indels   28;  Gaps    8;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      91219 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCACTGGAGTTGGGC 91278

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||||||||| ||||||||||||||  | |||| |||||
Db      91279 AACATGGCTTCTCCCCTTTCTAGGTCCTGTGGCAGCCATCTTGCTATCATTCGCCTTTGG 91338

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||    | ||||||||||||||||||||||||| ||||||||||||||
Db      91339 GCCCTGTATTTTTAACCCTCTTGTCAAATTTGTTTCCTCTAGGATCAAGGCCATCAAGCT 91398

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              |||||||||||||||||| ||||||||| ||||| |||||||||||| | |||||||
Db      91399 ACAGATGATCTTACAAATGGAACCCCAAATAAGCTCAGCTAACAACTTCTACCGAGGACC 91458

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||  |||||||||||||||||| |||||||||||| ||||||||||||||||||||
Db      91459 CCTGGAGTGACCCGCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 91518

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              |||||||||||||||||||||||||||||||||||||||||||||||||| || |||||||| ||
Db      91519 AACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATCGCCC 91578

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||||||||| || ||||||||||||||||||||||||
Db      91579 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGATGGGGGATTGAGAGGTGAAGCCAGCT 91638

Qy        420 GGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              || ||| ||||||| |||||||||||||||||||| |||||||||||| | |||||||
Db      91639 GGACTTCCTGGGTCGAGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAAA-GTTTGTAAA 91697

Qy        479 TGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTG 538
              |||||||||||||||||                  | ||||||||||||||  |||||
Db      91698 CGCACCAATCAGCACTCTG----------------TAAAAATGCACCAATCAGTGCTCTG 91741

Qy        539 TAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCAC 598
```

```
Db      91742 TAAAATGGACTAATCAGCAGAACGT-GGCGCGGCCAAATAAGGGAATAAAAGCTGGCCA- 91799

Qy        599 CCGAGCCAGCAGTGGCAACCCACTCGG------GTCCCCTTCCACACTGTGGAAGCTTTG 652
              |||||||||||||||||||||  |  ||      ||||||||||||||||||||||||||
Db      91800 CCGAGCCAGCAGTGGCAACCTGCCTGGGTCCCAGTCCCCTTCCACACTGTGGAAGCTTTG 91859

Qy        653 TTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTT 712
              ||||| |||| ||||||||||||||||||||||||||||||||| |||  ||||||||||
Db      91860 TTCTTCTGCTTTTCACAATAAATCTTGCTGCTGCTCATTCTTTGGGTCTGCACTACCTTT 91919

Qy        713 ATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCAC 771
              |||||||||||||||||| ||||||||| ||||||| |||| ||||||   |||||||
Db      91920 ATGAGCTGTAACACTCACTGTGAGGGTCTGCGGCTTCACTCCTCAAGTCAGTGAGACCAC 91979

Qy        772 GAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCA 831
              |||||||  | |||||||| |||||  ||||  |  ||||||||||||||||||||||
Db      91980 AAACCCACCAGGAGGAACAAACAACTCTGGATGCACCACCTTTAAGAGCTGTAACACTCA 92039

Qy        832 CTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAA 891
              |||||||| ||||  ||| ||||||||||||||  |||||||| |||||||  |||||||
Db      92040 CTGCGAAGGTCTGTGGCTCCACTCCTGAAGTCAGCAAGACCACGAACCCACTGGAAGGAA 92099

Qy        892 GAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGA 951
              |||||||||||||| ||||| ||||||||||||||||||||||  | |||||| |||
Db      92100 GAAACTCTGGACACATCTGAACATCTGAAGGAACAAACTCCAGACATGCTATCTTTAAGA 92159

Qy        952 GCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAAC 1011
              |||||||||||||    |||||| ||||||||||||||||||||| ||||||||||||
Db      92160 ACTGTAACACTCACCATGAGGGTCCGTGGCTTCATTCTTGAAGTCAGCGAGACCAAGAAC 92219

Qy       1012 CCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
              |||| ||||||||| |||||  |||||||  |||
Db      92220 CCACTGGAAGGAACCAATTCTGGACACAGACGGA 92253
```

RESULT 28
AC013406
LOCUS AC013406 190018 bp DNA linear PRI 21-APR-2005
DEFINITION Homo sapiens BAC clone RP11-341F20 from 2, complete sequence.
ACCESSION AC013406
VERSION AC013406.7 GI:10312292
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 190018)
AUTHORS Kalicki,J., Ureta,M. and Johnson,D.
TITLE The sequence of Homo sapiens BAC clone RP11-341F20
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 190018)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 190018)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 190018)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 190018)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (21-APR-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Sep 27, 2000 this sequence version replaced gi:9945236.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0341F20
----------.

NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-634D21; the clone sequenced to the right is RP11-219B1. Actual start of this clone is at base position 1 of RP11-341F20; actual end is at base position 190018 of RP11-341F20.

FEATURES Location/Qualifiers
source 1. .190018
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-341F20"
/clone_lib="RPCI-11"
misc_feature 72290. .72518
/note="CpG_island (%GC=62.0, o/e=0.86, #CpGs=19)"

ORIGIN

Query Match 38.0%; Score 761.2; DB 8; Length 190018;
Best Local Similarity 88.0%; Pred. No. 1.4e-191;
Matches 928; Conservative 0; Mismatches 98; Indels 28; Gaps 8;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      20888 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCACTGGAGTTGGGC 20947

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||||||||| ||||||||||||||  | |||| |||||
Db      20948 AACATGGCTTCTCCCCTTTCTAGGTCCTGTGGCAGCCATCTTGCTATCATTCGCCTTTGG 21007

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||    | |||||||||||||||||||||| ||||||||||||||
Db      21008 GCCCTGTATTTTTAACCCTCTTGTCAAATTTGTTTCCTCTAGGATCAAGGCCATCAAGCT 21067

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              |||||||||||||||||| |||||||||| |||||| |||||||||||| | |||||||
```

```
Db      21068 ACAGATGATCTTACAAATGGAACCCCAAATAAGCTCAGCTAACAACTTCTACCGAGGACC 21127

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||  |||||||||||||||||| ||||||||||| ||||||||||||||||||||
Db      21128 CCTGGAGTGACCCGCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 21187

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              |||||||||||||||||||||||||||||||||||||||||||||||| || ||||||| ||
Db      21188 AACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATCGCCC 21247

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| ||||||||||||||| || ||||||||||||||||||||||||
Db      21248 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGATGGGGGATTGAGAGGTGAAGCCAGCT 21307

Qy        420 GGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              || ||| |||||||  |||||||||||||||||||| |||||||||||| | |||||||
Db      21308 GGACTTCCTGGGTCGAGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAAA-GTTTGTAAA 21366

Qy        479 TGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTG 538
              ||||||||||||||||||                | ||||||||||||||| |||||
Db      21367 CGCACCAATCAGCACTCTG----------------TAAAAATGCACCAATCAGTGCTCTG 21410

Qy        539 TAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCAC 598
              |||||||||| |||||||||  | || |||| || |||||||||||| ||||| |||||||
Db      21411 TAAAATGGACTAATCAGCAGAACGT-GGCGCGGCCAAATAAGGGAATAAAAGCTGGCCA- 21468

Qy        599 CCGAGCCAGCAGTGGCAACCCACTCGG------GTCCCCTTCCACACTGTGGAAGCTTTG 652
              |||||||||||||||||||  |  ||      ||||||||||||||||||||||||||
Db      21469 CCGAGCCAGCAGTGGCAACCTGCCTGGGTCCCAGTCCCCTTCCACACTGTGGAAGCTTTG 21528

Qy        653 TTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTT 712
              |||||  |||| ||||||||||||||||||||||||||||||| |||  ||||||||||
Db      21529 TTCTTCTGCTTTTCACAATAAATCTTGCTGCTGCTCATTCTTTGGGTCTGCACTACCTTT 21588

Qy        713 ATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCAC 771
              |||||||||||||||||| ||||||||| ||||||| |||| ||||||   |||||||
Db      21589 ATGAGCTGTAACACTCACTGTGAGGGTCTGCGGCTTCACTCCTCAAGTCAGTGAGACCAC 21648

Qy        772 GAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCA 831
              |||||||  | |||||||| |||||  ||||  |  |||||||||||||||||||||
Db      21649 AAACCCACCAGGAGGAACAAACAACTCTGGATGCACCACCTTTAAGAGCTGTAACACTCA 21708

Qy        832 CTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAA 891
              |||||||| |||| ||| |||||||||||||||| ||||||| |||||||  |||||||
Db      21709 CTGCGAAGGTCTGTGGCTCCACTCCTGAAGTCAGCAAGACCACGAACCCACTGGAAGGAA 21768

Qy        892 GAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGA 951
              |||||||||||||||| ||||| |||||||||||||||||||||  | ||||| |||
Db      21769 GAAACTCTGGACACATCTGAACATCTGAAGGAACAAACTCCAGACATGCTATCTTTAAGA 21828

Qy        952 GCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAAC 1011
              |||||||||||||   |||||| |||||||||||||||||||| ||||||||||
Db      21829 ACTGTAACACTCACCATGAGGGTCCGTGGCTTCATTCTTGAAGTCAGCGAGACCAAGAAC 21888

Qy       1012 CCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
              |||| ||||||||| |||||  |||||||  |||
Db      21889 CCACTGGAAGGAACCAATTCTGGACACAGACGGA 21922
```

```
RESULT 29
AC098972
LOCUS       AC098972               145993 bp    DNA     linear   PRI 22-AUG-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-197J13, complete sequence.
ACCESSION   AC098972 AC031986
VERSION     AC098972.2  GI:22417369
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 145993)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 145993)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 145993)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Aug 22, 2002 this sequence version replaced gi:16756284.
            ---------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: BCM
            ---------- Project Information
                Center project name: chr-3
                Center clone name: RP11-197J13 (bc0295)
            ---------- Summary Statistics
                Sequencing vector: unknown; 45% of reads
                Sequencing vector: plasmid; 55% of reads
                Sequencing vector: plasmid; L08752; 0% of reads
                Chemistry: Dye-terminator ET; 80% of reads
                Chemistry: Dye-terminator Big Dye; 20% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 145920 bases at least Q40
                Consensus quality: 145984 bases at least Q30
                Consensus quality: 145993 bases at least Q20
                Insert size: 145993; sum-of-contigs
                Quality coverage: 11.4x in Q20 bases; sum-of-contigs
            ----------.
            Overlapping Sequences:
                5': RP11-655N11 (UWGC:bc0519) AC099057, 55267-bp overlap
                3': RP11-638F19 (UWGC:bc0505) AC107622, 53874-bp overlap
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
                   BglII                  NsiI                HindIII
```

SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
5583	5542	16518	16670	2251	2300
2067	2054	6048	6057	6382	6424
6916	7112	1968	1944	512	<800
1415	1397	8820	8839	449	<800
3539	3505	10950	10760	5313	5239
2898	2922	1125	1128	934	878
1716	1701	18381	18445	2061	2030
2858	2922	2937	3016	865	878
343	<800	150	<800	1557	1601
263	<800	462	<800	4158	4060
2502	2527	259	<800	683	<800
2325	2304	4527	4496	623	<800
2166	2161	380	<800	3684	3797
2301	2304	372	<800	3569	3610
3701	3698	545	<800	712	<800
1139	1148	81	<800	1265	1264
776	766	4438	4496	1030	994
1761	1701	5428	5485	2998	3112
5888	5931	160	<800	6492	6424
768	766	4957	4932	3155	3112
6502	6578	6821	6819	5926	5890
3345	3369	1025	1008	2538	2505
4233	4226	1307	1298	2324	2300
12029	11859	962	1008	562	<800
3455	3505	245	<800	2035	2030
603	<800	12092	11821	992	994
5243	5224	696	<800	891	878
2691	2683	769	757	412	<800
4386	4392	1278	1298	754	<800
3447	3505	514	<800	2731	2774
13647	13699	1636	1619	1281	1264
291	<800	9574	9578	77	<800
747	766	454	<800	1792	1868
9341	9385	1562	1619	3553	3478

333	<800	3718	3642	8284	8404
1437	1539	288	<800	8017	7944
864	880	7991	8076	260	<800
406	<800	2688	2762	566	<800
163	<800	1660	1619	1966	2030
214	<800	333	<800	532	<800
3057	2922	2678	2762	2414	2300
1347	1302	2764	2762	5570	5521
2330	2304	1352	1298	7795	7944
540	<800	1688	1619	779	765
347	<800	2091	2104	1617	1601
282	<800			1757	1789
424	<800			3387	3478
897	880			385	<800
1792	1792			2786	2774
185	<800			4095	3903
546	<800			3291	3238
2383	2417			2307	2300
1580	1701			1214	1264
4330	4226			1115	1114
4658	4577			2839	2935

Query Match 37.7%; Score 754.4; DB 8; Length 145993;
Best Local Similarity 83.7%; Pred. No. 9.3e-190;
Matches 934; Conservative 0; Mismatches 106; Indels 76; Gaps 4;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              | |||||||||||||||||||||||||||||||||||||||||||||||  |||||||||| |
Db     143358 ACAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAATGTCTGGAGTTGGAC 143417

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||| |||||||||||||||||||||||||| || | || |||||
Db     143418 AACATGGCTTCTCCCCTTCCTAGGTCCTGTGACAGCCATCTTGCTATTATTTGCCTTTGG 143477

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              | |||||||||||||||| |  |||| ||||||||||||||||||||||||||||||||
Db     143478 GACCTGTATTTTTAACCTCCT-TTCAAGTTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 143536

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| ||||||||||||||||||||||||||||| |  ||||||
Db     143537 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTACCAAGGACC 143596

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              | |||||| |||| |||||||||||| |||||||||||| ||||||||||||||||||
Db     143597 CTTGGACCAACCCACTGGCCCTTTCAGTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAG 143656

Qy        301 CACTGCAGGGCCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||||||||||| |  |||| ||||  |||||| |||||| || |||||||  ||
Db     143657 AACTGCAGGGCCCCTTCATAGCCCCCATCCCCCAGGAACTAGCTAGAGTGGTCATCACCC 143716

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              | |||||||||||| | |||||||||||||| ||||| ||| ||||||||||||||| |||
```

```
Db     143717 GATTCCCAACAGCAGTTAGGGTGTCCTGTTTAGAGGGAGGACTGAGAGGTGAAGCCGGCT 143776

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||  |||||||||||||||||||||| ||||||||||||||||||||||||
Db     143777 GGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTCGTGTCTAGCTAAAGGATTGTAAAT 143836

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT---- 535
              |||||||||||| |||||||||||||||||||| ||| |||||||||| ||||
Db     143837 GCACCAATCAGCGCTCTGTGTCTAGCTAAAGGATTGCAAACGCACCAATCATCACTCTGT 143896

Qy        536 ------------------------------------------------------------ 535

Db     143897 AAAAATGCACCAATCAGCACTCTGTAAAATGGGTCAATCAGCACTCTGTAAAATGGACCA 143956

Qy        536 ---------CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTA 586
                      |||||||||||||| |||||||||||||||||||| |||||||||||  ||
Db     143957 ATCAGCCCCCTGTAAAATGGACAAATCAGCAGGATGTGGGCGGGGCCAAATAAGGGAATA 144016

Qy        587 AAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAA 646
              ||| |||||||||| |||||||||||||||||  || ||||||| |||||| ||||||||
Db     144017 AAAGCTGGCCACCTGAGCCAGCAGTGGCAACCTGCTAGGGTCCCATTCCACGCTGTGGAA 144076

Qy        647 GCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACT 706
              ||||| ||| ||||||||||||||||||||||||||||||| |||||| ||||||||||
Db     144077 GCTTTATTCCTTTGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACT 144136

Qy        707 ACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACA 765
              |||||||||||||||||||||||||| || |||||| ||||||| ||||||||||   |
Db     144137 ACCTTTATGAGCTGTAACACTCACTGTGAAGGTCTGCGGCTTCACTCCTGAAGTCAGTGA 144196

Qy        766 GACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAA 825
              |||||||||||||| || ||||||||| |||||| ||    |   |||||||| |||||
Db     144197 GACCACGAACCCACCGGGAGGAACAAACAACTCCAGACACACCATCTTTAAGAGATGTAA 144256

Qy        826 CACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAG 885
              |||||||||||||| ||||  |||||||||||||||||||| |||||||| |||||||| |
Db     144257 CACTCACTGCGAAGGTCTGTGGCTTCACTCCTGAAGTCAGCGAGACCACGAACCCACCGG 144316

Qy        886 AAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCT 945
              |||||||||||||||||||| ||||| |||||||||||||||||||| |||||||||
Db     144317 AAGGAAGAAACTCTGGACACATCTGAACATCTGAAGGAACAAACTCCAGATACACCATCT 144376

Qy        946 TTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACC 1005
              || ||| ||||||||||||||   |||||| |||||||||||| |||||||||| ||| |
Db     144377 TTAAGAACTGTAACACTCACCATGAGGGTCCGTGGCTTCATTCGTGAAGTCAGCGAGATC 144436

Qy       1006 AAGAACCCACCGGAAGGAACAAATTCCAGACACAGT 1041
              |||||||||| ||||||||| |||||  |||||| |
Db     144437 AAGAACCCACTGGAAGGAACCAATTCTGGACACACT 144472
```

```
RESULT 30
AC107622
LOCUS       AC107622               158508 bp    DNA     linear   PRI 29-MAR-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-638F19, complete sequence.
ACCESSION   AC107622
VERSION     AC107622.2  GI:19807856
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 158508)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158508)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 158508)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Mar 29, 2002 this sequence version replaced gi:18308300.
            ------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            ------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-638F19 (bc0505)
            ------------ Summary Statistics
                Sequencing vector: plasmid; L08752; 6% of reads
                Sequencing vector: plasmid; 94% of reads
                Chemistry: Dye-terminator ET; 47% of reads
                Chemistry: Dye-terminator Big Dye; 53% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 158394 bases at least Q40
                Consensus quality: 158506 bases at least Q30
                Consensus quality: 158508 bases at least Q20
                Insert size: 158508; sum-of-contigs
                Quality coverage: 7.8x in Q20 bases; sum-of-contigs
            -------------.
            Overlapping Sequences:
                5':  RP11-197J13 (UWGC:bc0295) AC098972
                3':  RP11-669C19 (UWGC:bc0523) AC116096
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
```

HindIII		BglII		EcoRI	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
3990	3966	10712	10092	8696	8689
6382	6425	2067	2186	6	<800
512	<800	10737	10092	705	<800
449	<800	291	<800	705	<800

8252	8252	747	751	688	<800
779	<800	9340	9505	7499	7596
1616	1638	333	<800	7576	7596
1757	1765	1437	1405	589	<800
3387	3427	863	929	696	<800
385	<800	406	<800	205	<800
2786	2900	163	<800	12176	11807
4094	3966	214	<800	717	<800
3291	3244	3057	3217	645	<800
2307	2306	1347	1303	2251	2340
1214	1230	2330	2458	1982	1998
1115	1133	540	<800	3007	3036
2839	2900	347	<800	199	<800
2115	2051	282	<800	276	<800
321	<800	424	<800	1053	1036
3973	3966	897	929	486	<800
6441	6425	1792	1832	12153	11807
1899	1886	185	<800	258	<800
39	<800	546	<800	687	<800
4367	4285	2383	2562	3290	3295
2813	2900	1580	1579	442	<800
1063	1060	4330	4631	2289	2340
911	838	4658	4882	3697	3745
557	<800	1189	1175	2385	2519
5965	5999	4503	4702	2750	2885
2195	2306	3658	3798	241	<800
10103	10081	942	974	2017	2154
687	<800	10460	10092	5854	5595
1360	1371	6074	5775	1432	1494
791	808	2921	2899	860	859
3439	3427	1960	2113	1245	1192
1568	1550	2684	2618	121	<800
10038	10081	1647	1579	474	<800
2481	2581	228	<800	1552	1494
2858	2900	4411	4702	2761	2885
3849	3808	6484	6845	769	753

301	<800	5242	5139	2171	2340
1241	1230	2733	2899	10020	9872
1063	1060	769	<800	110	<800
1648	1638	3234	3217	6884	6817
1986	1886	3555	3472	3504	3531
131	<800	13420	12963	1242	1192
2540	2581	3394	3472	1371	1347
8828	8870	3073	3217	210	<800
941	934	6117	6509	291	<800
81	<800	1630	1579	5537	5595
5708	5610	14871	15751	675	<800
4925	4826			5542	5595
1890	1886			17062	17584
6562	6425			90	<800
2236	2306			6028	6011
463	<800			1561	1494
11675	11738			1164	1192

Query Match 37.7%; Score 754.4; DB 8; Length 158508;
Best Local Similarity 83.7%; Pred. No. 9.3e-190;
Matches 934; Conservative 0; Mismatches 106; Indels 76; Gaps 4;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              | |||||||||||||||||||||||||||||||||||||||||||  |||||||||| |
Db      51237 ACAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAATGTCTGGAGTTGGAC 51296

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||| |||||||||||||||||||||||| || | || |||||
Db      51297 AACATGGCTTCTCCCCTTCCTAGGTCCTGTGACAGCCATCTTGCTATTATTTGCCTTTGG 51356

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              | |||||||||||||||| |  |||| |||||||||||||||||||||||||||||||
Db      51357 GACCTGTATTTTTAACCTCCT-TTCAAGTTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 51415

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||||||||||| |  ||||||
Db      51416 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTACCAAGGACC 51475

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              | |||||| |||| |||||||||||| |||||||||||| ||||||||||||||||||
Db      51476 CTTGGACCAACCCACTGGCCCTTTCAGTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAG 51535

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||||||||||| |  |||| |||| |||||| |||||| || |||||||  ||
Db      51536 AACTGCAGGGCCCCTTCATAGCCCCCATCCCCCAGGAACTAGCTAGAGTGGTCATCACCC 51595

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              | |||||||||||| | |||||||||||||| ||||| ||| ||||||||||||||| |||
Db      51596 GATTCCCAACAGCAGTTAGGGTGTCCTGTTTAGAGGGAGGACTGAGAGGTGAAGCCGGCT 51655

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              ||||||||||||  |||||||||||||||||||| |||||||||||||||||||||||||
Db      51656 GGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTCGTGTCTAGCTAAAGGATTGTAAAT 51715

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT---- 535
              |||||||||||| |||||||||||||||||||||||| ||| |||||||||| ||||
```

```
Db      51716 GCACCAATCAGCGCTCTGTGTCTAGCTAAAGGATTGCAAACGCACCAATCATCACTCTGT 51775

Qy        536 ------------------------------------------------------------ 535

Db      51776 AAAAATGCACCAATCAGCACTCTGTAAAATGGGTCAATCAGCACTCTGTAAAATGGACCA 51835

Qy        536 ---------CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTA 586
                      |||||||||||||| |||||||||||||||||||||| |||||||||| ||
Db      51836 ATCAGCCCCCTGTAAAATGGACAAATCAGCAGGATGTGGGCGGGGCCAAATAAGGGAATA 51895

Qy        587 AAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAA 646
              ||| ||||||||| ||||||||||||||||||  || ||||||| |||||| ||||||||
Db      51896 AAAGCTGGCCACCTGAGCCAGCAGTGGCAACCTGCTAGGGTCCCATTCCACGCTGTGGAA 51955

Qy        647 GCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACT 706
              ||||| ||| ||||||||||||||||||||||||||||||| |||||| |||||||||
Db      51956 GCTTTATTCCTTTGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACT 52015

Qy        707 ACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACA 765
              |||||||||||||||||||||||||| || |||||| ||||||| ||||||||||   |
Db      52016 ACCTTTATGAGCTGTAACACTCACTGTGAAGGTCTGCGGCTTCACTCCTGAAGTCAGTGA 52075

Qy        766 GACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAA 825
              |||||||||||||| || ||||||||| |||||| ||     |   |||||||| |||||
Db      52076 GACCACGAACCCACCGGGAGGAACAAACAACTCCAGACACACCATCTTTAAGAGATGTAA 52135

Qy        826 CACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAG 885
              |||||||||||||| ||||  |||||||||||||||||| |||||||| |||||||| |
Db      52136 CACTCACTGCGAAGGTCTGTGGCTTCACTCCTGAAGTCAGCGAGACCACGAACCCACCGG 52195

Qy        886 AAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCT 945
              |||||||||||||||||||| ||||| |||||||||||||||||||| |||||||||
Db      52196 AAGGAAGAAACTCTGGACACATCTGAACATCTGAAGGAACAAACTCCAGATACACCATCT 52255

Qy        946 TTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACC 1005
              || ||| |||||||||||||   |||||| |||||||||||| |||||||||| ||| |
Db      52256 TTAAGAACTGTAACACTCACCATGAGGGTCCGTGGCTTCATTCGTGAAGTCAGCGAGATC 52315

Qy       1006 AAGAACCCACCGGAAGGAACAAATTCCAGACACAGT 1041
              |||||||||| ||||||||| |||||  |||||| |
Db      52316 AAGAACCCACTGGAAGGAACCAATTCTGGACACACT 52351


RESULT 31
BX571893/c
LOCUS       BX571893               94768 bp    DNA     linear   HTG 25-JUL-2003
DEFINITION  Homo sapiens chromosome 6 clone DASS-75O11, *** SEQUENCING IN
            PROGRESS ***, 4 unordered pieces.
ACCESSION   BX571893
VERSION     BX571893.1  GI:33284734
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 94768)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bSS75O11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 92859 bases at least Q40
            Consensus quality: 93559 bases at least Q30
            Consensus quality: 94137 bases at least Q20
            Insert size: 94468; sum-of-contigs
            Insert size: 99204; 1.9% error; agarose-fp
            Quality coverage: 4.88x in Q20 bases; sum-of-contigs Quality
            coverage: 4.81x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    31974: contig of 31974 bp in length
            *    31975    32074: gap of 100 bp
            *    32075    51357: contig of 19283 bp in length
            *    51358    51457: gap of 100 bp
            *    51458    83889: contig of 32432 bp in length
            *    83890    83989: gap of 100 bp
            *    83990    94768: contig of 10779 bp in length.
FEATURES             Location/Qualifiers
     source          1. .94768
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DASS-75O11"
                     /clone_lib="DNA-arts-BAC.1-SSTO.1"
     misc_feature    1. .31974
                     /note="assembly_fragment:00419
                     fragment_chain:1"
     misc_feature    32075. .51357
                     /note="assembly_fragment:00594
                     fragment_chain:1"
     misc_feature    51458. .83889
                     /note="assembly_fragment:00522
                     fragment_chain:1"
     misc_feature    83990. .94768
                     /note="assembly_fragment:00308
                     fragment_chain:1"
ORIGIN

  Query Match             37.6%;  Score 751.8;  DB 14;  Length 94768;
  Best Local Similarity   81.9%;  Pred. No. 4.6e-189;
  Matches  944;  Conservative    0;  Mismatches  117;  Indels   92;  Gaps    3;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      63182 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAAACTCTGGAGTTGGGC 63123

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||||||||||||||||||||| ||| || |||| |||||
Db      63122 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTACTATTACTCGCCTTTGG 63063

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||| |||||||||| | |||||||||||||| |||||||||| |||||| ||||||
Db      63062 GCCCTGTAGTTTTAACCTCCTTGTCAAATTTGTTTACTCTAGGATCAAGGCCAACAAGCT 63003

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| ||||||||| | |||||||||||||||| |||||||||| | ||||||
Db      63002 ACAGATGGTCTTACAAAGGGAACCCCAAATGAGCTCAGCTAACAACTTCTACCAAGGACC 62943

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||| |  ||  | ||| ||||||| | |||||||||||||| | ||||||||||||||||
Db      62942 CCTGAAATGAGTCACTGACCCTTTTACTGGCCTAAAGAGCTTCTCTCTGGAGGACACTAC 62883

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              |||||| |||||||||||||||||||||||||||||||||| || |||||| |||
Db      62882 AACTGCAAGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATTGCCC 62823

Qy        361 AAT-CCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
```

```
Db      62822 AATTCCCAACAGCAGTTGCCATGTCCTGTTTAGAGCGGGGATTGAGAGGTGAAGTCAGCT 62763

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479

Db      62762 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTCTGTTTAGCTAGAGGATTATAAAT 62703

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT---- 535

Db      62702 GCATCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAATGCACCAATCAGCACTGTGT 62643

Qy        536 ------------------------------------------------------------ 535

Db      62642 AAAAATGCACCAATCAGTGCTCTGTGTCTAGCTAAATGATTGTAAATGCACCAATCAACA 62583

Qy        536 -------------------------CTGTAAAATGGACCAATCAGCAGGATGTGGGCGG 569

Db      62582 CTCTGTAAAATGGACCAATCAGCACGCTGTAAAATGGACTAATCAGTAGGATGCGGGCAG 62523

Qy        570 GGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCC 629

Db      62522 GGCCAAATAAGGAAATAAAAGCTGGCCACCCAAGCCGGCAGTGGCAACTGGCTCAGGTCC 62463

Qy        630 CCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCA 689

Db      62462 ACTTCCACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCCGCTTA 62403

Qy        690 TTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTC 749

Db      62402 CTCTTTGGGTTCAGACTACCTTTATGAGCTGTAATACTCACCCTGAGGGTCTGCGGCTTC 62343

Qy        750 ATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCT 808

Db      62342 ATTCCTGAAGGCAGCAAGACCACGAACCCATCGGGAGGAACAAACAGCTCCAGACGTGCC 62283

Qy        809 GCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGA 868

Db      62282 ACCTTTAAGAGCTGTAACGCTCACTGCGAAGGTCTGCAGCTTCACTCCTGAAGTCAGCGA 62223

Qy        869 GACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAA 928

Db      62222 GACTACGAATCCACCAGAAGGAAGAAACTCCGGACACATCTGAACATCTGAAGGAACAAA 62163

Qy        929 CTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTC 988

Db      62162 CTCTGGACACACCATCTTTAAGAGCTGTAACATTCACCGCGAAGGTCTGTGGCTTCATTC 62103

Qy        989 TTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAAT 1048

Db      62102 TTGAAGTCAGTGAGACCAAGAACCCACCGGAAGGAATAAATTCCGGACACAGAAGGACTG 62043

Qy       1049 CTGTATTTTTGAT 1061

Db      62042 TCCCATTTTACAT 62030
```

```
RESULT 32
BX927235
LOCUS       BX927235               99242 bp    DNA     linear   PRI 16-FEB-2005
DEFINITION  Human DNA sequence from clone DAMA-25D9 on chromosome 6, complete
            sequence.
ACCESSION   BX927235
VERSION     BX927235.7  GI:59891245
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 99242)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 16, 2005 this sequence version replaced gi:45598013.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs constructed by
            the MHC Haplotype Consortium and collaborators. Further information
            can be found at
            http://www.sanger.ac.uk/HGP/Chr6/MHC
            DAMA-25D9 is from the DNA-Arts human BAC library MANN.1 VECTOR:
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .99242
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DAMA-25D9"
                     /clone_lib="DNA-Arts BAC library MANN.1"
ORIGIN
```

```
  Query Match             37.6%;  Score 751.8;  DB 8;  Length 99242;
  Best Local Similarity   81.9%;  Pred. No. 4.6e-189;
  Matches  944;  Conservative    0;  Mismatches  117;  Indels   92;  Gaps    3;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60

Db      37525 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAAACTCTGGAGTTGGGC 37584

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120

Db      37585 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTACTATTACTCGCCTTTGG 37644

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180

Db      37645 GCCCTGTAGTTTTAACCTCCTTGTCAAATTTGTTTACTCTAGGATCAAGGCCAACAAGCT 37704

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240

Db      37705 ACAGATGGTCTTACAAAGGGAACCCCAAATGAGCTCAGCTAACAACTTCTACCAAGGACC 37764

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300

Db      37765 CCTGAAATGAGTCACTGACCCTTTTACTGGCCTAAAGAGCTTCTCTCTGGAGGACACTAC 37824

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360

Db      37825 AACTGCAAGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATTGCCC 37884

Qy        361 AAT-CCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
```

```
Db      37885 AATTCCCAACAGCAGTTGCCATGTCCTGTTTAGAGCGGGGATTGAGAGGTGAAGTCAGCT 37944

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479

Db      37945 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTCTGTTTAGCTAGAGGATTATAAAT 38004

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT---- 535

Db      38005 GCATCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAATGCACCAATCAGCACTGTGT 38064

Qy        536 ------------------------------------------------------------ 535

Db      38065 AAAAATGCACCAATCAGTGCTCTGTGTCTAGCTAAATGATTGTAAATGCACCAATCAACA 38124

Qy        536 --------------------------CTGTAAAATGGACCAATCAGCAGGATGTGGGCGG 569

Db      38125 CTCTGTAAAATGGACCAATCAGCACGCTGTAAAATGGACTAATCAGTAGGATGCGGGCAG 38184

Qy        570 GGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCC 629

Db      38185 GGCCAAATAAGGAAATAAAAGCTGGCCACCCAAGCCGGCAGTGGCAACTGGCTCAGGTCC 38244

Qy        630 CCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCA 689

Db      38245 ACTTCCACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCCGCTTA 38304

Qy        690 TTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTC 749

Db      38305 CTCTTTGGGTTCAGACTACCTTTATGAGCTGTAATACTCACCCTGAGGGTCTGCGGCTTC 38364

Qy        750 ATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCT 808

Db      38365 ATTCCTGAAGGCAGCAAGACCACGAACCCATCGGGAGGAACAAACAGCTCCAGACGTGCC 38424

Qy        809 GCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGA 868

Db      38425 ACCTTTAAGAGCTGTAACGCTCACTGCGAAAGTCTGCAGCTTCACTCCTGAAGTCAGCGA 38484

Qy        869 GACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAA 928

Db      38485 GACTACGAATCCACCAGAAGGAAGAAACTCCGGACACATCTGAACATCTGAAGGAACAAA 38544

Qy        929 CTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTC 988

Db      38545 CTCTGGACACACCATCTTTAAGAGCTGTAACATTCACCGCGAAGGTCTGTGGCTTCATTC 38604

Qy        989 TTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAAT 1048

Db      38605 TTGAAGTCAGCGAGACCAAGAACCCACCGGAAGGAATAAATTCCGGACACAGAAGGACTG 38664

Qy       1049 CTGTATTTTTGAT 1061

Db      38665 TCCCATTTTACAT 38677
```

```
RESULT 33
CR788250
LOCUS       CR788250              104114 bp    DNA     linear   PRI 12-OCT-2004
DEFINITION  Human DNA sequence from clone DAMC-146E17 on chromosome 6, complete
            sequence.
ACCESSION   CR788250
VERSION     CR788250.5  GI:54110604
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 104114)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 12, 2004 this sequence version replaced gi:53145969.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs constructed by
            the MHC Haplotype Consortium and collaborators. Further information
            can be found at
            http://www.sanger.ac.uk/HGP/Chr6/MHC
            DAMC-146E17 is from the DNA-Arts.org BAC library MCF.1 VECTOR:
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .104114
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DAMC-146E17"
                     /clone_lib="DNA-Arts.org BAC library MCF.1"
ORIGIN

  Query Match             37.6%;  Score 751.8;  DB 8;  Length 104114;
  Best Local Similarity   81.9%;  Pred. No. 4.6e-189;
  Matches  944;  Conservative    0;  Mismatches  117;  Indels   92;  Gaps    3;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60

Db      74796 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAAACTCTGGAGTTGGGC 74855

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120

Db      74856 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTACTATTACTCGCCTTTGG 74915

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180

Db      74916 GCCCTGTAGTTTTAACCTCCTTGTCAAATTTGTTTACTCTAGGATCAAGGCCAACAAGCT 74975

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240

Db      74976 ACAGATGGTCTTACAAAGGGAACCCCAAATGAGCTCAGCTAACAACTTCTACCAAGGACC 75035

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300

Db      75036 CCTGAAATGAGTCACTGACCCTTTTACTGGCCTAAAGAGCTTCTCTCTGGAGGACACTAC 75095

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360

Db      75096 AACTGCAAGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATTGCCC 75155

Qy        361 AAT-CCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
```

```
Db      75156 AATTCCCAACAGCAGTTGCCATGTCCTGTTTAGAGCGGGGATTGAGAGGTGAAGTCAGCT 75215

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479

Db      75216 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTCTGTTTAGCTAGAGGATTATAAAT 75275

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT---- 535

Db      75276 GCATCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAATGCACCAATCAGCACTGTGT 75335

Qy        536 ------------------------------------------------------------ 535

Db      75336 AAAAATGCACCAATCAGTGCTCTGTGTCTAGCTAAATGATTGTAAATGCACCAATCAACA 75395

Qy        536 -------------------------CTGTAAAATGGACCAATCAGCAGGATGTGGGCGG 569

Db      75396 CTCTGTAAAATGGACCAATCAGCACGCTGTAAAATGGACTAATCAGTAGGATGCGGGCAG 75455

Qy        570 GGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCC 629

Db      75456 GGCCAAATAAGGAAATAAAAGCTGGCCACCCAAGCCGGCAGTGGCAACTGGCTCAGGTCC 75515

Qy        630 CCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCA 689

Db      75516 ACTTCCACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCCGCTTA 75575

Qy        690 TTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTC 749

Db      75576 CTCTTTGGGTTCAGACTACCTTTATGAGCTGTAATACTCACCCTGAGGGTCTGCGGCTTC 75635

Qy        750 ATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCT 808

Db      75636 ATTCCTGAAGGCAGCAAGACCACGAACCCATCGGGAGGAACAAACAGCTCCAGACGTGCC 75695

Qy        809 GCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGA 868

Db      75696 ACCTTTAAGAGCTGTAACGCTCACTGCGAAGGTCTGCAGCTTCACTCCTGAAGTCAGCGA 75755

Qy        869 GACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAA 928

Db      75756 GACTACGAATCCACCAGAAGGAAGAAACTCCGGACACATCTGAACATCTGAAGGAACAAA 75815

Qy        929 CTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTC 988

Db      75816 CTCTGGACACACCATCTTTAAGAGCTGTAACATTCACCGCGAAGGTCTGTGGCTTCATTC 75875

Qy        989 TTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAAT 1048

Db      75876 TTGAAGTCAGTGAGACCAAGAACCCACCGGAAGGAATAAATTCCGGACACAGAAGGACTG 75935

Qy       1049 CTGTATTTTTGAT 1061

Db      75936 TCCCATTTTACAT 75948
```

```
RESULT 34
BX120007
LOCUS       BX120007              109682 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone DASS-397D15 on chromosome 6, complete
            sequence.
ACCESSION   BX120007
VERSION     BX120007.9  GI:31335526
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 109682)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 2, 2003 this sequence version replaced gi:30024440.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs constructed by
            the MHC Haplotype Consortium and collaborators. Further information
            can be found at
            http://www.sanger.ac.uk/HGP/Chr6/MHC
            DASS-397D15 is from a DNA-arts SSTO human bac library VECTOR:
            pBeloBAC11
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .109682
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DASS-397D15"
                     /clone_lib="DNA-arts-BAC.1-SSTO.1"
     misc_feature    1
                     /note="Clone_left_end: DASS-397D15"
     misc_feature    12660
                     /note="Clone_right_end: DASS-380D18"
     misc_feature    13315. .13395
                     /note="Single clone region. Sequence generated from a
                     transposon library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    56811. .56841
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    109682
                     /note="Clone_right_end: DASS-397D15"
ORIGIN
```

```
  Query Match             37.6%;  Score 751.8;  DB 8;  Length 109682;
  Best Local Similarity   81.9%;  Pred. No. 4.6e-189;
  Matches  944;  Conservative    0;  Mismatches  117;  Indels   92;  Gaps    3;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60

Db      87451 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAAACTCTGGAGTTGGGC 87510

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120

Db      87511 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTACTATTACTCGCCTTTGG 87570

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180

Db      87571 GCCCTGTAGTTTTAACCTCCTTGTCAAATTTGTTTACTCTAGGATCAAGGCCAACAAGCT 87630

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240

Db      87631 ACAGATGGTCTTACAAAGGGAACCCCAAATGAGCTCAGCTAACAACTTCTACCAAGGACC 87690

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300

Db      87691 CCTGAAATGAGTCACTGACCCTTTTACTGGCCTAAAGAGCTTCTCTCTGGAGGACACTAC 87750
```

```
Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              |||||| |||||||||||||||||||||||||||||||||||||| || |||||| |||
Db      87751 AACTGCAAGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATTGCCC 87810

Qy        361 AAT-CCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              ||| |||||||||||  ||   |||||||||  ||| ||||||||||||||||||  |||||
Db      87811 AATTCCCAACAGCAGTTGCCATGTCCTGTTTAGAGCGGGGATTGAGAGGTGAAGTCAGCT 87870

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||||||||||||||||||||||||||||||||| ||| |||||| |||||| |||||
Db      87871 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTCTGTTTAGCTAGAGGATTATAAAT 87930

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT---- 535
              ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      87931 GCATCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAATGCACCAATCAGCACTGTGT 87990

Qy        536 ------------------------------------------------------------ 535

Db      87991 AAAAATGCACCAATCAGTGCTCTGTGTCTAGCTAAATGATTGTAAATGCACCAATCAACA 88050

Qy        536 -------------------------CTGTAAAATGGACCAATCAGCAGGATGTGGGCGG 569
                                        |||||||||||||| |||||| |||||| |||| |
Db      88051 CTCTGTAAAATGGACCAATCAGCACGCTGTAAAATGGACTAATCAGTAGGATGCGGGCAG 88110

Qy        570 GGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCC 629
              || |||||||||  | |||||  |||||||||| |||| ||||||||||    ||| |||||
Db      88111 GGCCAAATAAGGAAATAAAAGCTGGCCACCCAAGCCGGCAGTGGCAACTGGCTCAGGTCC 88170

Qy        630 CCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCA 689
              |||||||  |||||||||||||||||||  |||||||||||||||||||||||||| ||| |
Db      88171 ACTTCCACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCCGCTTA 88230

Qy        690 TTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTC 749
              |||||| || || |||||||||||||||||||| ||||||   |||||||||  ||||||
Db      88231 CTCTTTGGGTTCAGACTACCTTTATGAGCTGTAATACTCACCCTGAGGGTCTGCGGCTTC 88290

Qy        750 ATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCT 808
              |||||||||| || | |||||||||||||   || |||||||| | |||| || ||||
Db      88291 ATTCCTGAAGGCAGCAAGACCACGAACCCATCGGGAGGAACAAACAGCTCCAGACGTGCC 88350

Qy        809 GCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGA 868
              |||||||||||||||||| |||||||||||| |||||||||||||||||||||||||| ||
Db      88351 ACCTTTAAGAGCTGTAACGCTCACTGCGAAGGTCTGCAGCTTCACTCCTGAAGTCAGCGA 88410

Qy        869 GACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAA 928
              ||| || || |||||||||||||||||||||| |||||||| ||||| ||||||||||||||
Db      88411 GACTACGAATCCACCAGAAGGAAGAAACTCCGGACACATCTGAACATCTGAAGGAACAAA 88470

Qy        929 CTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTC 988
              |||  |||||||||||||| |||||||||||| ||||||| | ||||||||||||||||||
Db      88471 CTCTGGACACACCATCTTTAAGAGCTGTAACATTCACCGCGAAGGTCTGTGGCTTCATTC 88530

Qy        989 TTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAAT 1048
              ||||||||||  |||||||||||||||||||||||| ||||||| ||||||| ||||
Db      88531 TTGAAGTCAGTGAGACCAAGAACCCACCGGAAGGAATAAATTCCGGACACAGAAGGACTG 88590

Qy       1049 CTGTATTTTTGAT 1061
                  |||||  ||
Db      88591 TCCCATTTTACAT 88603


RESULT 35
BX571806/c
LOCUS       BX571806              197067 bp    DNA     linear   HTG 31-JUL-2003
DEFINITION  Homo sapiens chromosome 6 clone DASS-105O12, 23 unordered pieces.
ACCESSION   BX571806
VERSION     BX571806.2  GI:33386601
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 197067)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 31, 2003 this sequence version replaced gi:32997038.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bSS105012
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 187436 bases at least Q40
            Consensus quality: 190482 bases at least Q30
            Consensus quality: 192676 bases at least Q20
            Insert size: 194867; sum-of-contigs
            Insert size: 136251; 16.1% error; agarose-fp
            Quality coverage: 3.71x in Q20 bases; sum-of-contigs Quality
            coverage: 5.55x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17025: contig of 17025 bp in length
            *    17026    17125: gap of 100 bp
            *    17126    20038: contig of 2913 bp in length
            *    20039    20138: gap of 100 bp
            *    20139    30881: contig of 10743 bp in length
            *    30882    30981: gap of 100 bp
            *    30982    41727: contig of 10746 bp in length
            *    41728    41827: gap of 100 bp
            *    41828    47839: contig of 6012 bp in length
            *    47840    47939: gap of 100 bp
            *    47940    62377: contig of 14438 bp in length
            *    62378    62477: gap of 100 bp
            *    62478    65276: contig of 2799 bp in length
            *    65277    65376: gap of 100 bp
            *    65377    72991: contig of 7615 bp in length
            *    72992    73091: gap of 100 bp
            *    73092    79539: contig of 6448 bp in length
            *    79540    79639: gap of 100 bp
            *    79640    85615: contig of 5976 bp in length
            *    85616    85715: gap of 100 bp
            *    85716    98517: contig of 12802 bp in length
            *    98518    98617: gap of 100 bp
            *    98618   103829: contig of 5212 bp in length
            *   103830   103929: gap of 100 bp
            *   103930   119341: contig of 15412 bp in length
            *   119342   119441: gap of 100 bp
            *   119442   127313: contig of 7872 bp in length
            *   127314   127413: gap of 100 bp
            *   127414   145490: contig of 18077 bp in length
            *   145491   145590: gap of 100 bp
            *   145591   155206: contig of 9616 bp in length
            *   155207   155306: gap of 100 bp
            *   155307   157763: contig of 2457 bp in length
            *   157764   157863: gap of 100 bp
            *   157864   162040: contig of 4177 bp in length
            *   162041   162140: gap of 100 bp
            *   162141   168837: contig of 6697 bp in length
            *   168838   168937: gap of 100 bp
            *   168938   180653: contig of 11716 bp in length
```

```
          *   180654   180753: gap of 100 bp
          *   180754   187735: contig of 6982 bp in length
          *   187736   187835: gap of 100 bp
          *   187836   194764: contig of 6929 bp in length
          *   194765   194864: gap of 100 bp
          *   194865   197067: contig of 2203 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197067
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DASS-105O12"
                     /clone_lib="DNA-arts-BAC.1-SSTO.1"
     misc_feature    1. .17025
                     /note="assembly_fragment:00433
                     fragment_chain:1"
     misc_feature    17126. .20038
                     /note="assembly_fragment:00355
                     fragment_chain:1"
     misc_feature    20139. .30881
                     /note="assembly_fragment:00385
                     fragment_chain:1"
     misc_feature    30982. .41727
                     /note="assembly_fragment:00872
                     fragment_chain:1"
     misc_feature    41828. .47839
                     /note="assembly_fragment:00083
                     fragment_chain:2"
     misc_feature    47940. .62377
                     /note="assembly_fragment:01078
                     fragment_chain:2"
     misc_feature    62478. .65276
                     /note="assembly_fragment:00283
                     fragment_chain:3"
     misc_feature    65377. .72991
                     /note="assembly_fragment:00572
                     fragment_chain:3"
     misc_feature    73092. .79539
                     /note="assembly_fragment:00648
                     fragment_chain:4"
     misc_feature    79640. .85615
                     /note="assembly_fragment:00488
                     fragment_chain:4"
     misc_feature    85716. .98517
                     /note="assembly_fragment:00715
                     fragment_chain:5"
     misc_feature    98618. .103829
                     /note="assembly_fragment:01139
                     fragment_chain:5"
     misc_feature    103930. .119341
                     /note="assembly_fragment:00046.0"
     misc_feature    119442. .127313
                     /note="assembly_fragment:00049"
     misc_feature    127414. .145490
                     /note="assembly_fragment:00276"
     misc_feature    145591. .155206
                     /note="assembly_fragment:00381.0"
     misc_feature    155307. .157763
                     /note="assembly_fragment:00593"
     misc_feature    157864. .162040
                     /note="assembly_fragment:00684"
     misc_feature    162141. .168837
                     /note="assembly_fragment:00713.0"
     misc_feature    168938. .180653
                     /note="assembly_fragment:00882"
     misc_feature    180754. .187735
                     /note="assembly_fragment:00940"
     misc_feature    187836. .194764
                     /note="assembly_fragment:01037"
     misc_feature    194865. .197067
                     /note="assembly_fragment:01039"
ORIGIN
```

```
  Query Match           37.6%;  Score 751.8;  DB 14;  Length 197067;
  Best Local Similarity   81.9%;  Pred. No. 4.6e-189;
  Matches  944;  Conservative    0;  Mismatches  117;  Indels   92;  Gaps    3;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db      11498 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAAACTCTGGAGTTGGGC 11439

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||||||||||||||||||||||||| ||| || |||| |||||
Db      11438 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTACTATTACTCGCCTTTGG 11379

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||| |||||||||| | |||||||||||||| |||||||||| |||||| ||||||
Db      11378 GCCCTGTAGTTTTAACCTCCTTGTCAAATTTGTTTACTCTAGGATCAAGGCCAACAAGCT 11319

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| ||||||||| | |||||||||||||||| |||||||||||| | ||||||
Db      11318 ACAGATGGTCTTACAAAGGGAACCCCAAATGAGCTCAGCTAACAACTTCTACCAAGGACC 11259

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||| |  ||  | ||| |||||| | ||||||||||||| | ||||||||||||||||
Db      11258 CCTGAAATGAGTCACTGACCCTTTTACTGGCCTAAAGAGCTTCTCTCTGGAGGACACTAC 11199

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||| |||||||||||||||||||||||||||||||||| || |||||| |||
Db      11198 AACTGCAAGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATTGCCC 11139

Qy        361 AAT-CCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              ||| |||||||||||| ||   |||||||||| ||| ||||||||||||||||||| |||||
Db      11138 AATTCCCAACAGCAGTTGCCATGTCCTGTTTAGAGCGGGGATTGAGAGGTGAAGTCAGCT 11079

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||||||||||||||||||||||||||||||||||| ||| |||||| |||||
Db      11078 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTCTGTTTAGCTAGAGGATTATAAAT 11019

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT---- 535
              ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      11018 GCATCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAATGCACCAATCAGCACTGTGT 10959

Qy        536 ------------------------------------------------------------ 535

Db      10958 AAAAATGCACCAATCAGTGCTCTGTGTCTAGCTAAATGATTGTAAATGCACCAATCAACA 10899

Qy        536 --------------------------CTGTAAAATGGACCAATCAGCAGGATGTGGGCGG 569
                                         |||||||||||||| |||||| |||||| |||| |
Db      10898 CTCTGTAAAATGGACCAATCAGCACGCTGTAAAATGGACTAATCAGTAGGATGCGGGCAG 10839

Qy        570 GGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCC 629
              || ||||||||| | ||||| |||||||||| |||| ||||||||||   ||| |||||
Db      10838 GGCCAAATAAGGAAATAAAAGCTGGCCACCCAAGCCGGCAGTGGCAACTGGCTCAGGTCC 10779

Qy        630 CCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCA 689
              ||||||| |||||||||||||||||||  |||||||||||||||||||||||||| ||| |
Db      10778 ACTTCCACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCCGCTTA 10719

Qy        690 TTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTC 749
              |||||| || || |||||||||||||||||| ||||||   ||||||||| ||||||
Db      10718 CTCTTTGGGTTCAGACTACCTTTATGAGCTGTAATACTCACCCTGAGGGTCTGCGGCTTC 10659

Qy        750 ATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCT 808
              |||||||||| || | |||||||||||||  || |||||||| | |||| || ||||
Db      10658 ATTCCTGAAGGCAGCAAGACCACGAACCCATCGGGAGGAACAAACAGCTCCAGACGTGCC 10599

Qy        809 GCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGA 868
              |||||||||||||||| |||||||||||| ||||||||||||||||||||||||||| ||
Db      10598 ACCTTTAAGAGCTGTAACGCTCACTGCGAAGGTCTGCAGCTTCACTCCTGAAGTCAGCGA 10539

Qy        869 GACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAA 928
              ||| || || |||||||||||||||||||| ||||||| |||| ||||||||||||||||
Db      10538 GACTACGAATCCACCAGAAGGAAGAAACTCCGGACACATCTGAACATCTGAAGGAACAAA 10479
```

```
Qy        929 CTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTC 988
              ||| |||||||||||||| |||||||||||| ||||||| | ||||||||||||||||||
Db      10478 CTCTGGACACACCATCTTTAAGAGCTGTAACATTCACCGCGAAGGTCTGTGGCTTCATTC 10419

Qy        989 TTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAAT 1048
              |||||||||| |||||||||||||||||||||||| ||||||| ||||||| ||||
Db      10418 TTGAAGTCAGTGAGACCAAGAACCCACCGGAAGGAATAAATTCCGGACACAGAAGGACTG 10359

Qy       1049 CTGTATTTTTGAT 1061
                  |||||  ||
Db      10358 TCCCATTTTACAT 10346
```

RESULT 36
AC084277/c
LOCUS AC084277 150705 bp DNA linear PRI 13-MAY-2005
DEFINITION Homo sapiens BAC clone RP11-68O1 from 4, complete sequence.
ACCESSION AC084277
VERSION AC084277.3 GI:18042347
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150705)
AUTHORS Meyer,R. and Desai,A.
TITLE The sequence of Homo sapiens BAC clone RP11-68O1
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 150705)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 150705)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 150705)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 150705)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Jan 3, 2002 this sequence version replaced gi:17647044.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0068O01
----------.

NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-650N7; the clone sequenced to the right is RP11-769N3, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-68O1.

Data from AC083851 and AC021940 was used to finish this clone, AC084277. Polymorphisms have been identified between AC083851, AC021940 and AC084277.

FEATURES Location/Qualifiers
source 1. .150705
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-68O1"
/clone_lib="RPCI-11"

ORIGIN

Query Match 37.3%; Score 745.6; DB 8; Length 150705;
Best Local Similarity 84.1%; Pred. No. 2.1e-187;
Matches 940; Conservative 0; Mismatches 109; Indels 69; Gaps 6;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      65511 AAAGGCTTCCGAAGTCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 65452

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||||||  || ||||||||||||||| || || | || ||
Db      65451 AACATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTATTACTCACCTTCGG 65392

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||||||| |||| ||||||| |||||
Db      65391 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTA-GATCAAGGCCATGAAGCT 65333

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||  |||||||||||| |||||||||||||||||||||||||||||| | |||||||
Db      65332 ACAGATTGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTACCGAGGACC 65273

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||| ||| | |||||||||||| |||||||||||| |||||| ||||||||||||||
Db      65272 CCTGGACCAACCTGTTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTGTGGAGGACACTAC 65213

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              |||||||||||||||||  || ||||||||||||||||||||  ||  |||||| ||
Db      65212 AACTGCAGGGCCCCTTCTTTGCCACTATCCAGCAGGAAGTAGCTAGAGTAGTCATCACCC 65153

Qy        361 AA-TCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||||||| |||||||||||||||||||||||  |||
Db      65152 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCTGGCT 65093

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db      65092 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTCTCTCTAGCTAAAGGATTGTAAAT 65033

Qy        480 ------------------------------------------------------------ 479

Db      65032 GCACCAATCAGCACTGTGTCTAGCTAAAGGTTGGTAAATGCACCAATCAGCACTCTGTAA 64973
```

```
Qy        480 ----GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535
              ||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||
Db      64972 AAACGCACCAATCAGCACTCTGTGTCTAGCTAAAGGTTTGTAAAGGCACCAATCAGCACT 64913

Qy        536 CTGT-AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGG 594
              |||| |||||||||||||||||||||||||| |||| ||||||||| | ||||| || |
Db      64912 CTGTAAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAGGAAATAAAAGCTAG 64853

Qy        595 CCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA-CACTGTGGAAGCTTTGT 653
              ||||||||||||||||  ||||| |||||||||||||||||| ||||||||||||||||||
Db      64852 CCACCCGAGCCAGCAGCAGCAACACACTCGGGTCCCCTTCCACCACTGTGGAAGCTTTGT 64793

Qy        654 TCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTA 713
              |||||  |||||  |||||||||||||||||||| |||||| |||  ||||| |||||
Db      64792 TCTTTCACTCTTTGCAATAAATCTTGCTGCTGCTCACTCTTTGGGTCTGCACTATCTTTA 64733

Qy        714 TGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCACG 772
              |||||||||||||||| || | ||||||  |||||| ||||||||||    ||||||||
Db      64732 TGAGCTGTAACACTCACCGCAAAGGTCTGCAGCTTCACTCCTGAAGTCATTGAGACCACG 64673

Qy        773 AACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCAC 832
              |||||||||| |||||| || |||||| || | ||    || ||||||||||||||||||
Db      64672 AACCCACTGGGAGGAACGAACAACTCCGGACGCGCCAACTCTAAGAGCTGTAACACTCAC 64613

Qy        833 TGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAG 892
              || |||| ||||| |||| |||||||||||||||  ||||||| ||||||||||||||||||
Db      64612 TGTGAAGGTCTGCGGCTTTACTCCTGAAGTCAGCAAGACCACGAACCCACCAGAAGGAAG 64553

Qy        893 AAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAG 952
              ||||||  |||||| ||||| |||||||||| || ||||  |||||||||||||| |||
Db      64552 AAACTCCTGACACATCTGAACATCTGAAGGATCACACTCTGGACACACCATCTTTAAGAA 64493

Qy        953 CTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACC 1012
              || |||||||||||||| ||||||||  |||||||||||||||||||| ||||||||||||
Db      64492 CTATAACACTCACCGCGAGGGTCTGCAGCTTCATTCTTGAAGTCAGCGAGACCAAGAACC 64433

Qy       1013 CACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCT 1050
              |||| |||||||| |||||  ||||||| ||||   ||
Db      64432 CACCAGAAGGAACCAATTCTGGACACAGGAGGATCACT 64395
```

```
RESULT 37
AC145142
LOCUS       AC145142               184449 bp    DNA     linear   HTG 10-JUN-2003
DEFINITION  Homo sapiens chromosome 5 clone RP11-1398P22, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC145142
VERSION     AC145142.1  GI:31560141
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 184449)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184449)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 2756848
            Center clone name: RPCI-11_1398P22
            -------
            Summary Statistics
            Consensus quality: 184227 bases at least Q40
            Consensus quality: 184248 bases at least Q30
            Consensus quality: 184249 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 184249; sum-of-contigs estimation
            Quality coverage:  12.87 in Q20 bases; agarose-fp estimation
            Quality coverage:  12.22 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1188: contig of 1188 bp in length
            *     1189     1288: gap of unknown length
            *     1289    63133: contig of 61845 bp in length
            *    63134    63233: gap of unknown length
            *    63234   184449: contig of 121216 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184449
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-1398P22"
                     /clone_lib="RPCI human BAC library 11"
     gap             1189. .1288
                     /estimated_length=unknown
     gap             63134. .63233
                     /estimated_length=unknown
ORIGIN
```

```
  Query Match             37.3%;  Score 745.6;  DB 14;  Length 184449;
  Best Local Similarity   84.1%;  Pred. No. 2e-187;
  Matches  940;  Conservative    0;  Mismatches  109;  Indels   69;  Gaps    6;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     181095 AAAGGCTTCCGAAGTCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 181154

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||  || |||||||||||||| || || | || ||
Db     181155 AACATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTATTACTCACCTTCGG 181214

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||||||| |||| ||||||| |||||
Db     181215 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTA-GATCAAGGCCATGAAGCT 181273

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||  |||||||||| |||||||||||||||||||||||||||| | |||||||
Db     181274 ACAGATTGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTACCGAGGACC 181333

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||| ||| | |||||||||| |||||||||| |||||| ||||||||||||
Db     181334 CCTGGACCAACCTGTTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTGTGGAGGACACTAC 181393

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              |||||||||||||||||  || |||||||||||||||||| ||  |||||| ||
Db     181394 AACTGCAGGGCCCCTTCTTTGCCACTATCCAGCAGGAAGTAGCTAGAGTAGTCATCACCC 181453

Qy        361 AA-TCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||||||| |||||||||||||||||||||  |||
Db     181454 AATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCTGGCT 181513

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db     181514 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTCTCTCTAGCTAAAGGATTGTAAAT 181573

Qy        480 ------------------------------------------------------------ 479

Db     181574 GCACCAATCAGCACTGTGTCTAGCTAAAGGTTGGTAAATGCACCAATCAGCACTCTGTAA 181633
```

```
Qy        480 ----GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535
              ||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||
Db     181634 AAACGCACCAATCAGCACTCTGTGTCTAGCTAAAGGTTTGTAAAGGCACCAATCAGCACT 181693

Qy        536 CTGT-AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGG 594
              |||| |||||||||||||||||||||||||| |||| ||||||||| | ||||| || |
Db     181694 CTGTAAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAGGAAATAAAAGCTAG 181753

Qy        595 CCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA-CACTGTGGAAGCTTTGT 653
              ||||||||||||||   ||||| |||||||||||||||||| |||||||||||||||||
Db     181754 CCACCCGAGCCAGCAGCAGCAACACACTCGGGTCCCCTTCCACCACTGTGGAAGCTTTGT 181813

Qy        654 TCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTA 713
              |||||  |||||  |||||||||||||||||||||| |||||| |||  ||||| |||||
Db     181814 TCTTTCACTCTTTGCAATAAATCTTGCTGCTGCTCACTCTTTGGGTCTGCACTATCTTTA 181873

Qy        714 TGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCACG 772
              |||||||||||||||| || | ||||||  |||||| ||||||||||   ||||||||
Db     181874 TGAGCTGTAACACTCACCGCAAAGGTCTGCAGCTTCACTCCTGAAGTCATTGAGACCACG 181933

Qy        773 AACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCAC 832
              |||||||||| |||||| || |||||| || | ||   || ||||||||||||||||||
Db     181934 AACCCACTGGGAGGAACGAACAACTCCGGACGCGCCAACTCTAAGAGCTGTAACACTCAC 181993

Qy        833 TGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAG 892
              || |||| ||||| |||| |||||||||||||  ||||||| ||||||||||||||||||
Db     181994 TGTGAAGGTCTGCGGCTTTACTCCTGAAGTCAGCAAGACCACGAACCCACCAGAAGGAAG 182053

Qy        893 AAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAG 952
              ||||||  |||||| ||||| |||||||||| || ||||  |||||||||||||| |||
Db     182054 AAACTCCTGACACATCTGAACATCTGAAGGATCACACTCTGGACACACCATCTTTAAGAA 182113

Qy        953 CTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACC 1012
              || |||||||||||| ||||||||  |||||||||||||||||| ||||||||||||
Db     182114 CTATAACACTCACCGCGAGGGTCTGCAGCTTCATTCTTGAAGTCAGCGAGACCAAGAACC 182173

Qy       1013 CACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCT 1050
              |||| |||||||| |||||  ||||||| ||||   ||
Db     182174 CACCAGAAGGAACCAATTCTGGACACAGGAGGATCACT 182211
```

```
RESULT 38
AC074133/c
LOCUS       AC074133               75852 bp    DNA     linear   PRI 28-FEB-2002
DEFINITION  Homo sapiens chromosome 5 clone CTD-2278B20, complete sequence.
ACCESSION   AC074133
VERSION     AC074133.5  GI:18997252
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 75852)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 75852)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 75852)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 75852)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Feb 28, 2002 this sequence version replaced gi:18266630.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99% of Sequence;
            Estimated Total Number of Errors is 0.6.
            NOTE:  This insert is not the entire sequence of the clone (entire
            sequence is 112kb). It is clipped at the overlap with AC008883. The
            number of bases overlapped is 13829.
FEATURES             Location/Qualifiers
     source          1. .75852
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2278B20"
ORIGIN
```

```
  Query Match             37.1%;  Score 742.2;  DB 8;  Length 75852;
  Best Local Similarity   85.6%;  Pred. No. 1.7e-186;
  Matches  936;  Conservative    0;  Mismatches  108;  Indels   49;  Gaps    8;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||||||||||||||||||||||||||||| ||| |||| |||||| |
Db      36812 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAACTCTTATACTAACTTCTGCAGTTGGAC 36753

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              || |||||||||||||| ||||||| ||| |||||||||||||| || | || |||||
Db      36752 AACCTGGCTTCTCCCCTTTTTAGGTCCCGTGGCAGCCATCTTGCTATTACTTGCCTTTGG 36693

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||| |||||||||||||| | ||  ||||||||||||||||||||| |||||||||| |
Db      36692 GCCATGTATTTTTAACCTCCTTGT-TAATTTGTTTCCTCTAGGATCAAGGCCATCAAGTT 36634

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||| |||||||||||||| | ||||| |
Db      36633 ACAGATGGTCTTACAAATGGAACCCCAAATGAGTTCAACTAACAACTTCTACCGAGGAAC 36574

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||| |||| |||||||||||||| ||||||| |||| ||||||||||||||||||||
Db      36573 CCTGGACTGACCTGCTGGCCCTTTCACTGGCCTAGAGAGTTCCCCTCTGGAGGACACTAC 36514

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCA-TCGCC 359
              | || |||||||||||| ||||||| |||||||||||||||||||| |||||||| | ||
Db      36513 AATTGTAGGGCCCCTTCATCACCCC-ATCCAGCAGGAAGTAGCTAGAGCGGTCACTGCCC 36455

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||| |||||||||| ||||||||||||||||||||||||||||
Db      36454 AATTCCCAACAGCAGTTGGGTTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCAGCT 36395

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTT---------------------------- 454
              |||||||||||| |||||||||||||||||||| ||
Db      36394 GGGCTTCTGGGTCGGGTGGGGACTTGGAGAACCTTTCTGTCTAGCTAAAGGATTGTAAGC 36335

Qy        455 ----------------TGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT 498
                              |||||||||||||||| ||||||| |||||||||||| |||||
Db      36334 ACACCAATCAGCGCTCTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATCAGCATTCTGT 36275

Qy        499 GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAG 558
              |||||||||||| |||||||  ||||||||| ||||||||||||||||||||||||||||
Db      36274 GTCTAGCTAAAGGTTTGTAAACACACCAATCAACACTCTGTAAAATGGACCAATCAGCAG 36215

Qy        559 GATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACC 618
              ||  |||| | | |||||||||||| ||||| || |||||| ||||||||||||||||||
Db      36214 GACATGGG-TGAGCCAAATAAGGGAATAAAAGCTAGCCACCTGAGCCAGCAGTGGCAACC 36156

Qy        619 CACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTT 678
              |||| |||||||||||  ||||||||||||||||||||| ||||||||||||||||||||
Db      36155 TACTCAGGTCCCCTTCCATGCTGTGGAAGCTTTGTTCTTTTACTCTTCACAATAAATCTT 36096

Qy        679 GCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCT--GTAACACTCACTGCGAG 736
              |||||||||| |||||| |||  ||||||||||||||||   |||||||||||| ||
```

```
Db      36095 GCTGCTGCTCACTCTTTGGGTCAGCACTACCTTTATGAGCTACATAACACTCACTGTGAA 36036

Qy        737 GGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAA 795
              |||||||||||||| |||||||||| | |||||||||||||||| | ||||||||| ||
Db      36035 GGTCTGTGGCTTCACTCCTGAAGTCAGCGAGACCACGAACCCACTAGGAGGAACAAACAA 35976

Qy        796 CTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTC 855
              |||  || | ||  |||||||| |||||||||||||||||||| ||||| |||||||||
Db      35975 CTCTGGACGCGCCACCTTTAAGGGCTGTAACACTCACTGCGAAGGTCTGCGGCTTCACTC 35916

Qy        856 CTGAAGTC-AGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATA 914
              |||||||| ||  ||||||| |||||||  ||||||| |||||||||||||||  |||| |
Db      35915 CTGAAGTCAAGCAAGACCACGAACCCACTGGAAGGAAAAAACTCTGGACACATTTGAACA 35856

Qy        915 TCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGT 974
              |||||||||||||||  |||||||||||||| ||| ||||||||||||| |  |||||
Db      35855 TCTGAAGGAACAAACTCTGGACACACCATCTTTAAGAACTGTAACACTCACTGTGAGGGT 35796

Qy        975 CTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAG 1034
              | ||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||
Db      35795 CCGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCAGAAGGAACCAATTCCAG 35736

Qy       1035 ACACAGTAGGAAA 1047
              ||||| ||  |||
Db      35735 ACACAATACTAAA 35723


RESULT 39
AC091977/c
LOCUS       AC091977               183494 bp    DNA     linear   PRI 15-DEC-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-528L24, complete sequence.
ACCESSION   AC091977
VERSION     AC091977.3  GI:17861016
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 183494)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183494)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 183494)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 15, 2001 this sequence version replaced gi:15290435.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.7% of Sequence;
            Estimated Total Number of Errors is 0.3.
FEATURES             Location/Qualifiers
     source          1. .183494
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-528L24"
ORIGIN

  Query Match           37.1%;  Score 742.2;  DB 8;  Length 183494;
  Best Local Similarity   83.8%;  Pred. No. 1.6e-186;
  Matches  918;  Conservative    0;  Mismatches  128;  Indels   49;  Gaps    5;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||| |||||||||||||||||||||||||||||||||||| |||||
Db      35692 AAAGGCTTCTGAAATCACACAATGCCTTTCAAACTCTTATACCAACCTCTGGAATTGGGT 35633

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||| |||||||||||||||||||||||||||||| || |  | ||  |
Db      35632 GACATGGCTTCTCCTCTTTCTAGGTCCTGTGACAGCCATCTTGCTATTACTTTCCTTCAG 35573

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||| |||||  ||||||||||||||||||||||||||||||||||||||
Db      35572 GCCCTGTATTTTTGACCTCCCTGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 35513

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||    |||||||||| | ||||| |
Db      35512 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTC----AACAACTTCTACCGAGGATC 35457

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||| |||| |||||||||||||| |||||| ||||| |||||||||||||||||||
Db      35456 CCTGGACTGACCTGCTGGCCCTTTCACTGGCCTGAAGAGTTCCCCTCTGGAGGACACTAC 35397

Qy        301 CACTGCAGGGCCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||||||| |||||||||||| |||||||||||||||||| ||| ||||||  ||
Db      35396 AACTGCAGGGCCCTTTCTTCACCCCTGTCCAGCAGGAAGTAGCTAGAGCAGTCATCACCC 35337

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| ||| |||||||||| |||||||||||||||||||||||||||||
Db      35336 AATTCCCAACAGCAGTTGGAGTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCAGCT 35277

Qy        420 GGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              || ||| ||||||   |||||||| |||||||||| |||||||||| ||||||||||
Db      35276 GGACTTCCTGGGTTGAGTGGGGACATGGAGAACTTTTCTGTCTAGCTAGAGGATTGTAAA 35217

Qy        479 TGCACCAATCAG----------------------------------------CACTCT 496
              |||||||||||                                        |   |
Db      35216 TGCACCAATCAGTGCCTAGCTAAAGGATTGTAAATGCACCAATCCGCACTCTGTAAAAAT 35157

Qy        497 GTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGC 556
              |   | | |   |   |  ||||| |||||||||  ||||||||||||||||| |||||
Db      35156 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGGGCTCTGTAAAATGGACCAGTCAGC 35097

Qy        557 AGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAA 616
              |||||||||||||| |||||||| || ||||| |||||||||  ||||  |||  | ||
Db      35096 AGGATGTGGGCGGGGCCAAATAAGAGAATAAAAGCTGGCCACCCCAGCCCACAGCAGTAA 35037

Qy        617 CCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATC 676
              ||||||| |||||||||||||  |||||||||||||||||||||||||| ||||| ||||||
Db      35036 CCCACTCAGGTCCCCTTCCATGCTGTGGAAGCTTTGTTCTTTTGCTCCTCACAGTAAATC 34977

Qy        677 TTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAG 736
              |||||||||||||| |||||| |||  |||||||||||||||||||||||||||| ||
Db      34976 TTGCTGCTGCTCACTCTTTGGGTCTGCACTACCTTTATGAGCTGTAACACTCCCTGTGAA 34917

Qy        737 GGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAA 795
              ||||||  |||||| |||||||||||| | |||||||  |||||||||  ||||| ||| ||
Db      34916 GGTCTGCAGCTTCACTCCTGAAGTCAGCAAGACCACAAACCCACTGGGAGGAATAAACAA 34857

Qy        796 CTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTC 855
              ||||  |   |||  |||||||||||||||||||||||||||||| |||||  |||||||||
Db      34856 CTCCGGCCATGCCAACTTTAAGAGCTGTAACACTCACTGCGAAGGTCTGCGGCTTCACTC 34797

Qy        856 CTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATAT 915
              |||||||||||||||| | |||||||  |||||||||||||||||| ||| || || ||
Db      34796 CTGAAGTCAGTGAGACCCCGAACCCACTGGAAGGAAGAAACTCTGGAGACATCTAAACAT 34737

Qy        916 CTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTC 975
              |||||||||||||||||||||||||||||||| ||| |||||||||||| |  ||||||
Db      34736 CTGAAGGAACAAACTCCAGACACACCATCTTTAAGAACTGTAACACTCACTGTGAGGGTC 34677

Qy        976 TGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGA 1035
              || |||||||||||||||||| |||||||||||| ||| |||||||| |||||| ||
Db      34676 TGCGGCTTCATTCTTGAAGTCAGCGAGACCAAGAACCTACCAGAAGGAACCAATTCCGGA 34617

Qy       1036 CACAGTAGGAAATCT 1050
```

```
              |||| |   | || |
Db      34616 CACAATTTAACATGT 34602


RESULT 40
AC104697
LOCUS       AC104697               159797 bp    DNA     linear   PRI 21-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-454B3 from 2, complete sequence.
ACCESSION   AC104697 AC024403
VERSION     AC104697.2  GI:18855149
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 159797)
  AUTHORS   Cotton,M. and Radionenko,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-454B3
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 159797)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 159797)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 159797)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 21, 2002 this sequence version replaced gi:17933873.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0454B03
                Drafting Center: WIBR
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-337B18, 2000 bp overlap;
            the clone sequenced to the right is RP11-241F6, 2000 bp overlap.
            Actual start of this clone is at base position 167881 of
            RP11-337B18; actual end is at base position 2890 of RP11-241F6.

            Polymorphisms have been identified between AC092632, AC096560, and
            AC024403.

            The sequence of AC024403 has been incorporated into AC104697.
FEATURES             Location/Qualifiers
     source          1. .159797
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-454B3"
                     /clone_lib="RPCI-11"
     misc_feature    55931. .56157
                     /note="CpG_island (%GC=62.6, o/e=0.91, #CpGs=20)"
     misc_feature    59015. .59288
                     /note="CpG_island (%GC=59.9, o/e=0.74, #CpGs=20)"
ORIGIN

  Query Match             37.0%;  Score 741.2;  DB 8;  Length 159797;
  Best Local Similarity   85.1%;  Pred. No. 3.1e-186;
  Matches  921;  Conservative    0;  Mismatches  128;  Indels   33;  Gaps    7;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||||| |||||||||||||||||||||||||||| || |||
Db      29127 AAAGGCTTCTGAAATCAGACAATACCTTTCAAACTCTTATACCAACCTCTGGACTTCGGC 29186

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||| || |||||||||||||||  |||||||||||||||| || | || || ||
Db      29187 AACATGGATTTTCCCCTTTCTAGGTCCCATGACAGCCATCTTGCTATTACTTGCCTTCGG 29246

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||| ||||||||| || ||||||||||||||
Db      29247 GCCCTGTATTTTTAACCTCCTTGTCAAATTTG-TTCCTCTAGAATTGAGGCCATCAAGCT 29305

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| ||||||||||||||||||||||||||||| | |||||||
Db      29306 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTACCGAGGACC 29365

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||| ||  |||| ||| |||| ||| ||| || | || ||||||||||||||||
Db      29366 CCTGGACCAACTTGCTGACCCCTTCACTGGTCTACAGTGTTCTCCTCTGGAGGACACTAC 29425

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              ||| |||| ||||||||| ||||||||||||||||||||||||||||| ||||||  ||
Db      29426 AACTACAGGTCCCCTTCTTTGCCCCTATCCAGCAGGAAGTAGCTACAGCAGTCATCACCC 29485

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||| |||||||||| ||| ||||||||||||||||||||||||
Db      29486 AATTCCCAACAGCAGTTGGGCTGTCCTGTTTAGAGTGGGGATTGAGAGGTGAAGCCAGCT 29545

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29546 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 29605

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT 539
              |||||||||| |||||||                  |  |||  ||||||||||||||||||
Db      29606 GCACCAATCAACACTCTG----------------TAAAAACACACCAATCAGCACTCTGT 29649

Qy        540 AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACC 599
              |||||||||||||||||||  ||||| ||| |||||||||| ||||| |||||||||
Db      29650 AAAATGGACCAATCAGCAGGACATGGGCAGGGCCAAATAAGGGAATAAAAGCTGGCCACC 29709

Qy        600 CGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTT 659
              |||||||||| |||||| || ||||||||||||||||  ||||||||||||||||||||
Db      29710 CGAGCCAGCAGCAGCAACCTGCTGGGGTCCCCTTCCACAC--TGGAAGCTTTGTTCTTTT 29767
```

```
Qy        660 GCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCT 719
              ||||||||||||||||||||||||||||||||||||| |||||| |||| |||||||||| |||||
Db      29768 GCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCGCACTACCTTTAGGAGCT 29827

Qy        720 GTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCA 778
              ||||||||| ||| | ||  |||||||||| || | |||||| | ||||||||||||||
Db      29828 ATAACACTCATTGCAAAGGGTTGTGGCTTCACTCTTCAAGTCAGCAAGACCACGAACCCA 29887

Qy        779 CTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAA 838
              |||| |||||||           |||| |  ||||||||||| |||||||||||| |||
Db      29888 CTGGGAGGAACA----------ATGTACCACCTTTAAGAGCTCTAACACTCACTGAGAA 29936

Qy        839 GCTCTGCAGCTTCACTCCTGAAGTC-AGTGAGACCACAAACCCACCAGAAGGAAGAAACT 897
              | |||||||||||||||||||||| |||||||||| ||||||| ||| |||||||||
Db      29937 GGTCTGCAGCTTCACTCCTGAAGTCAAGTGAGACCACGAACCCACTGGAATGAAGAAACT 29996

Qy        898 CTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTA 957
              |  |||||  ||||  |||||||||||||||||  | ||||||||||| ||| |||||
Db      29997 CCAGACACGTCTGACCATCTGAAGGAACAAACTCTGGGCACACCATCTTTAAGAACTGTA 30056

Qy        958 ACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCG 1017
              ||| ||||||| ||||||||  |||||||||||||||||||| ||||||||||||||||
Db      30057 ACAGTCACCGCGAGGGTCTGCAGCTTCATTCTTGAAGTCAGCGAGACCAAGAACCCACCG 30116

Qy       1018 GAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGT 1077
              |||||||| |||||  |||| | ||  | || |        | |   ||   |  || ||
Db      30117 GAAGGAACCAATTCTGGACATATTAATACATGTAAGAACATTTTAGATGTTGTAAAGTGT 30176

Qy       1078 TA 1079
              ||
Db      30177 TA 30178
```

RESULT 41
AC034238/c
LOCUS AC034238 117583 bp DNA linear PRI 31-OCT-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2313F11, complete sequence.
ACCESSION AC034238
VERSION AC034238.4 GI:16554348
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 117583)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 117583)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 117583)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Oct 31, 2001 this sequence version replaced gi:13677055.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.2.
FEATURES Location/Qualifiers
source 1. .117583
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2313F11"

ORIGIN

Query Match 37.0%; Score 740.6; DB 8; Length 117583;
Best Local Similarity 83.7%; Pred. No. 4.4e-186;
Matches 917; Conservative 0; Mismatches 129; Indels 49; Gaps 5;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||| |||||
Db     104788 AAAGGCTTCTGAAATCACACAATGCCTTTCAAACTCTTATACCAACCTCTGGAATTGGGT 104729

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||| |||||||||||||||||||||||||||||| || |  | ||  |
Db     104728 GACATGGCTTCTCCTCTTTCTAGGTCCTGTGACAGCCATCTTGCTATTACTTTCCTTCAG 104669

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||| |||||   ||||||||||||||||||||||||||||||||||||
Db     104668 GCCCTGTATTTTTGACCTCCCTGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 104609

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| ||||||||||||||    |||||||||| | ||||| |
Db     104608 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTC----AACAACTTCTACCGAGGATC 104553

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||| |||| |||||||||||| |||||| ||||| ||||||||||||||||||||
Db     104552 CCTGGACTGACCTGCTGGCCCTTTCACTGGCCTGAAGAGTTCCCCTCTGGAGGACACTAC 104493

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||||||| |||||||||||| |||||||||||||||||||| ||| ||||||  ||
Db     104492 AACTGCAGGGCCCTTTCTTCACCCCTGTCCAGCAGGAAGTAGCTAGAGCAGTCATCACCC 104433

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| ||| |||||||||| ||||||||||||||||||||||||||||||
Db     104432 AATTCCCAACAGCAGTTGGAGTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCAGCT 104373

Qy        420 GGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              || ||| ||||||   |||||||| |||||||||||| |||||||||| ||||||||||
Db     104372 GGACTTCCTGGGTTGAGTGGGGACATGGAGAACTTTTCTGTCTAGCTAGAGGATTGTAAA 104313

Qy        479 TGCACCAATCAG----------------------------------------CACTCT 496
              ||||||||||||                                        |   |
Db     104312 TGCACCAATCAGTGCCTAGCTAAAGGATTGTAAATGCACCAATCCGCACTCTGTAAAAAT 104253

Qy        497 GTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGC 556
              |  | | |   |  |  ||||| |||||||||  |||||||||||||||||| |||||
Db     104252 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGGGCTCTGTAAAATGGACCAGTCAGC 104193

Qy        557 AGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAA 616
              |||||||||| |||| |||||||| || ||||| |||||||||| ||||  |||  | ||
Db     104192 AGGATGTGGGTGGGGCCAAATAAGAGAATAAAAGCTGGCCACCCCAGCCCACAGCAGTAA 104133

Qy        617 CCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATC 676
              ||||||| |||||||||||||  |||||||||||||||||||||||||| ||||| ||||||
Db     104132 CCCACTCAGGTCCCCTTCCATGCTGTGGAAGCTTTGTTCTTTTGCTCCTCACAGTAAATC 104073

Qy        677 TTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAG 736
              |||||||||||| |||||| |||  |||||||||||||||||||||||| ||| ||
Db     104072 TTGCTGCTGCTCACTCTTTGGGTCTGCACTACCTTTATGAGCTGTAACACTCCCTGTGAA 104013

Qy        737 GGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAA 795
              |||||| |||||| |||||||||| | ||||||| |||||||||| ||||| ||| ||
Db     104012 GGTCTGCAGCTTCACTCCTGAAGTCAGCAAGACCACAAACCCACTGGGAGGAATAAACAA 103953

Qy        796 CTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTC 855
              |||| |   |||  |||||||||||||||||||||||||||||| ||||| |||||||||
Db     103952 CTCCGGTCATGCCAACTTTAAGAGCTGTAACACTCACTGCGAAGGTCTGCGGCTTCACTC 103893

Qy        856 CTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATAT 915
              |||||||||||||||| | |||||||  |||||||||||||||| ||| || || ||
Db     103892 CTGAAGTCAGTGAGACCCCGAACCCACTGGAAGGAAGAAACTCTGGAGACATCTAAACAT 103833

Qy        916 CTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTC 975
              |||||||||||||||||||||||||||||||| ||| |||||||||||| |  ||||||
Db     103832 CTGAAGGAACAAACTCCAGACACACCATCTTTAAGAACTGTAACACTCACTGTGAGGGTC 103773
```

```
Qy        976 TGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGA 1035
              || |||||||||||||||||||| |||||||||||| ||| |||||||| |||||| ||
Db     103772 TGCGGCTTCATTCTTGAAGTCAGCGAGACCAAGAACCTACCAGAAGGAACCAATTCCGGA 103713

Qy       1036 CACAGTAGGAAATCT 1050
              |||| |   | || |
Db     103712 CACAATTTAACATGT 103698


RESULT 42
AC091845/c
LOCUS       AC091845               133322 bp    DNA     linear   HTG 09-JUN-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2210H11, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC091845
VERSION     AC091845.1  GI:14333781
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 133322)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133322)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 708368
            Center clone name: CITB-H1_2210H11
            -------
            Summary Statistics
            Consensus quality: 129273 bases at least Q40
            Consensus quality: 131004 bases at least Q30
            Consensus quality: 131261 bases at least Q20
            Estimated insert size: 134580; agarose-fp estimation
            Estimated insert size: 132922; sum-of-contigs estimation
            Quality coverage:  7.34 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.43 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1253: contig of 1253 bp in length
            *     1254     1353: gap of unknown length
            *     1354     5163: contig of 3810 bp in length
            *     5164     5263: gap of unknown length
            *     5264     8261: contig of 2998 bp in length
            *     8262     8361: gap of unknown length
            *     8362    35118: contig of 26757 bp in length
            *    35119    35218: gap of unknown length
            *    35219   133322: contig of 98104 bp in length.
FEATURES            Location/Qualifiers
     source         1. .133322
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="CTD-2210H11"
                    /clone_lib="CalTech human BAC library D"
     gap            1254. .1353
                    /estimated_length=unknown
     gap            5164. .5263
                    /estimated_length=unknown
     gap            8262. .8361
                    /estimated_length=unknown
     gap            35119. .35218
                    /estimated_length=unknown
ORIGIN

  Query Match           37.0%;  Score 740.6;  DB 14;  Length 133322;
  Best Local Similarity  83.7%;  Pred. No. 4.4e-186;
  Matches  917;  Conservative    0;  Mismatches  129;  Indels   49;  Gaps    5;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||| ||||||||||||||||||||||||||||||||||| |||||
Db      26895 AAAGGCTTCTGAAATCACACAATGCCTTTCAAACTCTTATACCAACCTCTGGAATTGGGT 26836

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||| ||||||||||||||||||||||||||||||||| || |  | ||  |
Db      26835 GACATGGCTTCTCCTCTTTCTAGGTCCTGTGACAGCCATCTTGCTATTACTTTCCTTCAG 26776

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||| |||||  ||||||||||||||||||||||||||||||||||||||
Db      26775 GCCCTGTATTTTTGACCTCCCTGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 26716

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||    |||||||||| | ||||| |
Db      26715 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTC----AACAACTTCTACCGAGGATC 26660

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||| |||| |||||||||||||| |||||| ||||| ||||||||||||||||||
Db      26659 CCTGGACTGACCTGCTGGCCCTTTCACTGGCCTGAAGAGTTCCCCTCTGGAGGACACTAC 26600

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||||||| |||||||||||| |||||||||||||||||| ||| ||||||  ||
Db      26599 AACTGCAGGGCCCTTTCTTCACCCCTGTCCAGCAGGAAGTAGCTAGAGCAGTCATCACCC 26540

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| ||| |||||||||| ||||||||||||||||||||||||||||
Db      26539 AATTCCCAACAGCAGTTGGAGTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCAGCT 26480

Qy        420 GGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              || ||| ||||||   |||||||| |||||||||||| |||||||||| ||||||||||
Db      26479 GGACTTCCTGGGTTGAGTGGGGACATGGAGAACTTTTCTGTCTAGCTAGAGGATTGTAAA 26420

Qy        479 TGCACCAATCAG----------------------------------------CACTCT 496
              ||||||||||||                                        |   |
Db      26419 TGCACCAATCAGTGCCTAGCTAAAGGATTGTAAATGCACCAATCCGCACTCTGTAAAAAT 26360

Qy        497 GTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGC 556
              |   | | |   |   |   ||||| |||||||||  |||||||||||||||| |||||
Db      26359 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGGGCTCTGTAAAATGGACCAGTCAGC 26300

Qy        557 AGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAA 616
              |||||||||| |||| |||||||| || ||||| |||||||||| ||||  |||  | ||
Db      26299 AGGATGTGGGTGGGGCCAAATAAGAGAATAAAAGCTGGCCACCCCAGCCCACAGCAGTAA 26240

Qy        617 CCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATC 676
              ||||||| |||||||||||  |||||||||||||||||||||||||||| ||||| ||||||
Db      26239 CCCACTCAGGTCCCCTTCCATGCTGTGGAAGCTTTGTTCTTTTGCTCCTCACAGTAAATC 26180

Qy        677 TTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAG 736
              |||||||||||||| |||||| |||  |||||||||||||||||||||||||||| ||| ||
Db      26179 TTGCTGCTGCTCACTCTTTGGGTCTGCACTACCTTTATGAGCTGTAACACTCCCTGTGAA 26120

Qy        737 GGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAA 795
              ||||||  |||||| |||||||||||| | |||||||  |||||||||| ||||| ||| ||
Db      26119 GGTCTGCAGCTTCACTCCTGAAGTCAGCAAGACCACAAACCCACTGGGAGGAATAAACAA 26060

Qy        796 CTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTC 855
              |||| |   |||   |||||||||||||||||||||||||||| ||||| |||||||||
```

```
Db      26059 CTCCGGTCATGCCAACTTTAAGAGCTGTAACACTCACTGCGAAGGTCTGCGGCTTCACTC 26000

Qy        856 CTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATAT 915
              |||||||||||||||| | |||||||  |||||||||||||||||| ||| || || ||
Db      25999 CTGAAGTCAGTGAGACCCCGAACCCACTGGAAGGAAGAAACTCTGGAGACATCTAAACAT 25940

Qy        916 CTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTC 975
              |||||||||||||||||||||||||||||| ||| |||||||||||| |  ||||||
Db      25939 CTGAAGGAACAAACTCCAGACACACCATCTTTAAGAACTGTAACACTCACTGTGAGGGTC 25880

Qy        976 TGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGA 1035
              || |||||||||||||||||||| |||||||||||| ||| |||||||| |||||| ||
Db      25879 TGCGGCTTCATTCTTGAAGTCAGCGAGACCAAGAACCTACCAGAAGGAACCAATTCCGGA 25820

Qy       1036 CACAGTAGGAAATCT 1050
              |||| |   | || |
Db      25819 CACAATTTAACATGT 25805


RESULT 43
AL645568
LOCUS       AL645568               136031 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-296O14 on chromosome 1 Contains
            a pseudogene similar to part of ribosomal protein L26 (RPL26), a
            novel pseudogene and a novel gene, complete sequence.
ACCESSION   AL645568
VERSION     AL645568.9  GI:18072559
KEYWORDS    HTG; RPL26.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 136031)
  AUTHORS   Matthews,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 6, 2002 this sequence version replaced gi:17977921.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-296O14 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted; all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .136031
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-296O14"
                     /clone_lib="RPCI-11.2"
     misc_feature    1
                     /note="Clone_left_end: RP11-296O14"
     gene            11587. .11902
                     /locus_tag="RP11-296O14.1-001"
                     /pseudo
     CDS             11587. .11902
                     /locus_tag="RP11-296O14.1-001"
                     /note="match: proteins: Sw:P12749 Sw:P47832 Sw:Q02877
                     Sw:Q9UNX3 Tr:AAH17360 Tr:BAC21653 Tr:BAC56365 Tr:Q7SXA1
                     Tr:Q862N4 Tr:Q90YU2"
                     /pseudo
                     /codon_start=1
                     /product="pseudogene similar to part of ribosomal protein
                     L26 (RPL26)"
     gene            complement(22296. .23631)
                     /locus_tag="RP11-296O14.2-001"
                     /pseudo
     CDS             complement(22296. .23631)
                     /locus_tag="RP11-296O14.2-001"
                     /note="match: proteins: Tr:BAC97863 Tr:Q14704 Tr:Q810F2
                     Tr:Q865V0 Tr:Q86XR8 Tr:Q8CEE0 Tr:Q8IXP0 Tr:Q9BVF9
                     Tr:Q9D5W5"
                     /pseudo
                     /codon_start=1
                     /product="novel pseudogene"
     misc_feature    34722. .34863
                     /note="Sequence from overlapping clone RP11-192B4
                     (AL360002). Assembly confirmed by restriction digest."
     misc_feature    60074
                     /note="Clone_right_end: RP11-192B4"
     gene            complement(join(77691. .78520,120276. .121260))
                     /locus_tag="RP11-296O14.3-001"
     mRNA            complement(join(77691. .78520,120276. .121260))
                     /locus_tag="RP11-296O14.3-001"
                     /product="novel transcript"
                     /note="match: ESTs: Em:AA743066.1 Em:AA806822.1
                     Em:AA812123.1 Em:AI655890.1 Em:BX117627.1
                     match: cDNAs: Em:AK130460.1"
     polyA_site      complement(77691)
                     /locus_tag="RP11-296O14.3-001"
     polyA_signal    complement(77713. .77718)
                     /locus_tag="RP11-296O14.3-001"
     gene            complement(join(77972. .78520,100361. .100648,
                     101817. .101995))
                     /locus_tag="RP11-296O14.3-002"
     mRNA            complement(join(77972. .78520,100361. .100648,
                     101817. .101995))
                     /locus_tag="RP11-296O14.3-002"
                     /product="novel transcript"
                     /note="match: ESTs: Em:AL521783.2 Em:BQ222905.1
                     Em:BX345040.1"
     misc_feature    134032
                     /note="Clone_left_end: RP3-436N22"
ORIGIN

  Query Match             37.0%;  Score 740.4;  DB 8;  Length 136031;
  Best Local Similarity   82.0%;  Pred. No. 5e-186;
  Matches  945;  Conservative    0;  Mismatches  131;  Indels   76;  Gaps    5;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||||||||||| |||||||||| |||||||| ||||||||||
Db      41251 AAAGGCTTCTGAAATCAGACAATGCCTCTCAAACTCTTACACCAACCTGTGGAGTTGGGC 41310

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||||  ||||||||||||||||| || | |  |||||
Db      41311 AACATGGCTTCTCCCCTTTCTAGGTCCCATGACAGCCATCTTGCTATTACTTGTCTTTGG 41370

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||| |||||||||||| | ||||||| ||||||||| |||||| |||||| ||||||
Db      41371 GCCCTGAATTTTTAACCTCCTTGTCAAATGTGTTTCCTCCAGGATCAAGGCCACCAAGCT 41430

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
```

```
Db      41431 ACAGACGGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTACCAAGGACC 41490

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300

Db      41491 CCTGAACCAACCCATTGGCCTTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 41550

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359

Db      41551 AACTGCAGGGCTTCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCAGTCATTGCCC 41610

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419

Db      41611 AATTCCCAACAGCAGCTGGGGTGTCCTGTTTAGAGGGGGGATTGACAGGTGAAGCCGGCT 41670

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGT---- 475

Db      41671 GGGCTTCTGGGTCGGGTGGGGACTTGAAGAACTTTTCTGTCTAGCTAAAGGATTGTAAAC 41730

Qy        476 ------------------------------------------------------------ 475

Db      41731 ACACCAATCAGCGCTCTGTGTCTAGCTAAAGTTTGTAAATGCACCAAGCAGCACTCTGTA 41790

Qy        476 -AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCAC 534

Db      41791 AAAATGCACCAATCAGTGCTCTGTGTCTAGCTAAAGGTTTGTAAGTGCACCAATCAGTGC 41850

Qy        535 TCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGG 594

Db      41851 TCTGTGAAATGGACCAATCAGCAGGACGTGGGCAGGGTGAAATAAAGGAATAAAAGCTGG 41910

Qy        595 CCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTT 654

Db      41911 CCACCTGAGCCAGCAGCAGCAACCTGCTTGGGTCCCCTTCCACACTGTGGAAGTTTTGTT 41970

Qy        655 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTAT 714

Db      41971 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACTACCTTTAT 42030

Qy        715 GAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGA 773

Db      42031 GAGCTGTAACACTCACTGCGAAGGTCTGCAGCTTCACTCCTGAAGTCAGCAAGACCACAA 42090

Qy        774 ACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACT 833

Db      42091 ACCCACCGGGAGGAACAAACAACTCCAGACGCACCACCTTTAAGAGCTGTAACACTCACT 42150

Qy        834 GCGAAGCTCTGCAGCTTCACTCCTGAAGTC-AGTGAGACCACAAACCCACCAGAAGGAAG 892

Db      42151 GCAAAGGTCTGTGGCTTCACTCCTGAAATCAAGCTAGACCACGAACCCACCAGAATGAAG 42210

Qy        893 AAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAG 952

Db      42211 AAACTCTGGACAC--------ATCTGAAGGAACACAGTCCGGACACACCATCTTTAGGAA 42262

Qy        953 CTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACC 1012

Db      42263 CTCTAACACTCACCGTGAGGGTCCACAGCTTCATTCTTGAAGTCAGCAAGACCAAGAAAC 42322

Qy       1013 CACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCC 1072

Db      42323 CACCAGAAGGAACCAATTCCGGACACATTAGGAGAAATACCTAATGTAGATGACGGGTTG 42382

Qy       1073 AGGGTTACTCCA 1084

Db      42383 ATGAGTGCAGCA 42394


RESULT 44
AL591074
LOCUS       AL591074               203076 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone XXyac-60D10 on chromosome 20 Contains
            the CST5 gene for cystatin D, the CST2 gene for cystatin SA, the
            CST1 gene for cystatin SN and the CSTP2 gene for cystatin
            pseudogene 2, complete sequence.
ACCESSION   AL591074
VERSION     AL591074.10  GI:15858969
KEYWORDS    HTG; CST1; CST2; CST5; CSTP2; cystatin D; cystatin SA; cystatin SN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 203076)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 2, 2001 this sequence version replaced gi:15705266.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            XXyac-60D10 is from a ICRF YAC Library
            VECTOR: pYAC4
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .203076
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="XXyac-60D10"
                     /clone_lib="Human-ICRF-YAC"
     misc_feature    100
                     /note="Clone_right_end: RP1-234M6"
     gene            complement(1587. .1673)
                     /locus_tag="AL591074.2-002"
                     /pseudo
     CDS             complement(1587. .1673)
                     /locus_tag="AL591074.2-002"
                     /note="match: proteins: AAH21225 O19092 O19093 P01036
                     P01037 P09228"
                     /pseudo
                     /codon_start=1
     misc_feature    1919. .2301
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived form a clone PCR."
     misc_feature    14801. .15094
                     /note="Single clone region. Sequence from reads from a
                     short library derived from a clone PCR."
     misc_feature    15422. .15653
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a clone PCR."
     gene            complement(37829. .41204)
                     /gene="CST1"
                     /locus_tag="AL591074.1-001"
     mRNA            complement(join(37829. .38166,39283. .39396,40906. .41204))
```

```
                     /gene="CST1"
                     /locus_tag="AL591074.1-001"
                     /note="match: ESTs: BF689725 BF689815 BG287350 BG290010
                     BG290406 BG291293 BG403032 BG403934 BG765263
                     match: cDNAs: BC021225.1 J03870.1 X54667.1"
     polyA_site      complement(37829)
                     /gene="CST1"
                     /locus_tag="AL591074.1-001"
     CDS             complement(join(38083. .38166,39283. .39396,40906. .41133))
                     /gene="CST1"
                     /locus_tag="AL591074.1-001"
                     /standard_name="OTTHUMP00000030444"
                     /note="match: proteins: CAC07196 P01035 P01036 P01037"
                     /codon_start=1
                     /protein_id="CAC94783.1"
                     /db_xref="GI:16116524"
                     /db_xref="Genew:2473"
                     /db_xref="GOA:P01037"
                     /db_xref="HSSP:1G96"
                     /db_xref="UniProt/Swiss-Prot:P01037"
                     /translation="MAQYLSTLLLLLATLAVALAWSPKEEDRIIPGGIYNADLNDEWV
                     QRALHFAISEYNKATKDDYYRRPLRVLRARQQTVGGVNYFFDVEVGRTICTKSQPNLD
                     TCAFHEQPELQKKQLCSFEIYEVPWENRRSLVKSRCQES"
     misc_feature    73978. .74211
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a clone PCR."
     misc_feature    90914. .91427
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a clone PCR."
     gene            complement(114039. .116942)
                     /gene="CST2"
                     /locus_tag="AL591074.3-001"
     mRNA            complement(join(114039. .114370,115477. .115590,
                     116700. .116942))
                     /gene="CST2"
                     /locus_tag="AL591074.3-001"
                     /note="match: ESTs: AA523539 AI521546 BG700542 BG706328
                     BG708038 BI488523 BI549132 BI551519 BI601985
                     match: cDNAs: AF311741.1 BC013083.1 U52028 X05607
                     X05607.1"
     polyA_site      complement(114039)
                     /gene="CST2"
                     /locus_tag="AL591074.3-001"
     polyA_signal    complement(114049. .114054)
                     /gene="CST2"
                     /locus_tag="AL591074.3-001"
     CDS             complement(join(114287. .114370,115477. .115590,
                     116700. .116927))
                     /gene="CST2"
                     /locus_tag="AL591074.3-001"
                     /standard_name="OTTHUMP00000030445"
                     /note="match: proteins: AAK11570 P09228"
                     /codon_start=1
                     /protein_id="CAC94784.1"
                     /db_xref="GI:16116525"
                     /db_xref="Genew:2474"
                     /db_xref="GOA:P09228"
                     /db_xref="HSSP:1G96"
                     /db_xref="InterPro:IPR000010"
                     /db_xref="UniProt/Swiss-Prot:P09228"
                     /translation="MAWPLCTLLLLLATQAVALAWSPQEEDRIIEGGIYDADLNDERV
                     QRALHFVISEYNKATEDEYYRRLLRVLRAREQIVGGVNYFFDIEVGRTICTKSQPNLD
                     TCAFHEQPELQKKQLCSFQIYEVPWEDRMSLVNSRCQEA"
     gene            complement(166423. .169947)
                     /gene="CST5"
                     /locus_tag="AL591074.5-001"
     mRNA            complement(join(166423. .166538,167772. .167885,
                     169713. .169947))
                     /gene="CST5"
                     /locus_tag="AL591074.5-001"
                     /note="match: cDNAs: X70377"
     CDS             complement(join(166455. .166538,167772. .167885,
                     169713. .169943))
                     /gene="CST5"
                     /locus_tag="AL591074.5-001"
                     /standard_name="OTTHUMP00000030448"
                     /note="match: proteins: P28325"
                     /codon_start=1
                     /protein_id="CAI23644.1"
                     /db_xref="GI:56205430"
                     /db_xref="InterPro:IPR000010"
                     /translation="MMWPMHTPLLLLTALMVAVAGSASAQSRTLAGGIHATDLNDKSV
                     QCALDFAISEYNKVINKDEYYSRPLQVMAAYQQIVGGVNYYFNVKFGRTTCTKSQPNL
                     DNCPFNDQPKLKEEEFCSFQINEVPWEDKISILNYKCRKV"
     misc_feature    181333. .181442
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a clone PCR."
     misc_feature    202977
                     /note="Clone_left_end: RP11-1P13"
ORIGIN

  Query Match             37.0%;  Score 739.8;  DB 8;  Length 203076;
  Best Local Similarity   85.7%;  Pred. No. 7.2e-186;
  Matches  914;  Conservative    0;  Mismatches  122;  Indels   31;  Gaps    7;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||| | |||||||||||||| ||  |||||||| |||||||||||||||||||||||| ||
Db      31970 AAACGATTCTGAAATCAGATAACACCTTTCAAGCTCTTATACCAACCTCTGGAGTTGAGC 32029

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||||||||||||||||||||| |  || |||||||||||||| ||   || || ||
Db      32030 AACATGGCTTCTCCCCTTTCTAGGTACCCTGGCAGCCATCTTGCTATTACCTGCCTTCGG 32089

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||||| | |||||||||||||||||||||||| |||||| ||||||
Db      32090 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGGATTGAGGCCACCAAGCT 32149

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              |||||| ||||||| ||| || |||||||||||||||||||||||||| |  ||||||
Db      32150 TCAGATGGTCTTACATATGGAAACCCAAATGAGCTCAACTAACAACTTCTACCAAGGACC 32209

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||  |||| |||||||||||| |||||||||||| || |||||||||| |||||
Db      32210 CCTGGACTAACCCACTGGCCCTTTCACTGGCCTAAAGAGTTCACCTCTGGAGGATACTAC 32269

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||||||||||||||||  |||||||||||||||||||||||| || |||||||
Db      32270 AACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATCATTG 32329

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||||||| |||| |||||||| ||||||||||||||
Db      32330 AACTCCCAACAGCAGTTGGGGTGTCCTGTTT-GAGGAGGGATTGACAGGTGAAGCCAGCT 32388

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              |||||||||||||||||||||||||||||||||||||| ||| |||||||||||| ||||||
Db      32389 AGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTATGTGTAGCTAAAGGATGGTAAAT 32448

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACC-------------- 525
              |||||||||||||||||||||||||||||||| ||||||| |||||
Db      32449 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGTTTGTAAAGGCACCAATCAGCACTCTGT 32508

Qy        526 ----------AATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCA 574
                         |||||||||||||||||||||||||||||||||||||| |||| ||||| ||
Db      32509 AAAAACAGACTAATCAGCACTCTGTAAAATGGACCAATCAGCAGGTTGTGGGGGGGGCCA 32568

Qy        575 AATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTC 634
              |||||||| ||||| |||||||| || |||||||||||||||||||| ||||||||||
Db      32569 AATAAGGGAATAAAAGCTGGCCA-CCTAGCCAGCAGTGGCAACCCACTTGGGTCCCCTTC 32627

Qy        635 CACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTT 694
              ||| |||||||||||||||||  |||||||||||||||||||||||||||||||| ||||
Db      32628 CACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCTGCTCACTCTT 32687

Qy        695 TGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCC 754
              || |||| |||||||||||||||||||| |||||  ||| ||||||  |||||| |||
Db      32688 TGGGTCCTCACTACCTTTATGAGCTGTAACCCTCACCACGAAGGTCTGCAGCTTCACTCC 32747
```

```
Qy        755 TGAAGTCA-ACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813
              ||||||||   ||| ||| |||||||  | |||||  |  |||||| ||  | |  || |
Db      32748 TGAAGTCAGTGAGAGCACAAACCCACCAGGAGGAATGAGCAACTCCAGACATACCACC-T 32806

Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873
              |||||||||||||||||||| |||| ||||  |||||||| ||||||||||  ||||||
Db      32807 TAAGAGCTGTAACACTCACTGTGAAGGTCTGTGGCTTCACTTCTGAAGTCAGCAAGACCA 32866

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933
              | |||||||||||||||||||||  |||| |  |||| ||||||||||| |||||
Db      32867 CGAACCCACCAGAAGGAAGAAACTCCAGACATATATGAACATCTGAAGGAAC-AACTCTG 32925

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
              ||||||||||||| ||| |||||||||||||| |||||| |  |||| |||||||||
Db      32926 GACACACCATCTTTAAGAACTGTAACACTCACCGCCAGGGTCCGCAGCTTTATTCTTGAA 32985

Qy        994 GTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAG 1040
              ||| || || |||||||||||||| |||||||| |||||  |||||||
Db      32986 GTCGGCGAGGCCAAGAACCCACCAGAAGGAACCAATTCAGGACACAG 33032
```

RESULT 45
AC026307/c
LOCUS AC026307 166357 bp DNA linear PRI 01-FEB-2001
DEFINITION Homo sapiens 12 BAC RP11-753N8 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC026307
VERSION AC026307.16 GI:12621201
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166357)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166357)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 166357)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 166357)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jan 31, 2001 this sequence version replaced gi:12585011.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

---------- Summary Statistics ----------

Contig length:	166357
Phrap values in estimate:	165586
Average error rate (BCM-Phrap estimate):	7.20476e-06
Fraction of Phrap values less than 40 :	0.00725907
Number of consensus changing edits:	184
Number of N's in consensus :	0

---------- Consensus changing edits ----------

Position	Original+Context	Edited+Context
5852	gttagggttc(n)ttttttttt	gttagggttc(t)ttttttttt
11194	ataaatagcc(n)naatcagagg	ataaatagcc(a)gaatcagagg
11195	taaatagccn(n)aatcagaggg	taaatagcca(g)aatcagaggg
22657	cacttctata(n)tggatattca	cacttctata(t)tggatattca
22704	ctatacttag(n)tcttcatctt	ctatacttag(t)tcttcatctt
29549	tgcaacctca(n)gaagcctaaa	tgcaacctca(g)gaagcctaaa
39699	tctattcttc(n)gaaggcaact	tctattcttc(t)gaaggcaact
42431	tttggtaatt(n)nnnnnnnaan	tttggtaatt(c)atatgagaat
42432	ttggtaattn(n)nnnnnnaann	ttggtaattc(a)tatgagaata
42433	tggtaattnn(n)nnnnaannn	tggtaattca(t)atgagaatag
42434	ggtaattnnn(n)nnnnaannna	ggtaattcat(a)tgagaataga
42435	gtaattnnnn(n)nnnaannnaa	gtaattcata(t)gagaatagaa
42436	taattnnnnn(n)nnaannnaan	taattcatat(g)agaatagaat
42437	aattnnnnnn(n)naannnaann	aattcatatg(a)gaatagaatg
42438	attnnnnnnn(n)aannnaannt	attcatatga(g)aatagaatgt
42441	nnnnnnnnaa(n)nnaannttnn	catatgagaa(t)agaatgttct
42442	nnnnnnnaan(n)naannttnnn	atatgagaat(a)gaatgttcta
42443	nnnnnnaann(n)aannttnnnn	tatgagaata(g)aatgttctag
42446	nnnaannnaa(n)nttnnnnttt	gagaatagaa(t)gttctagttt
42447	nnaannnaan(n)ttnnnnttt	agaatagaat(g)ttctagtttt
42450	annnaanntt(n)nnnttttttnt	atagaatgtt(c)tagtttttct
42451	nnnaannttn(n)nnttttttntt	tagaatgttc(t)agtttttctt
42452	nnaannttnn(n)ntttttnttn	agaatgttct(a)gttttcttg
42453	naannttnnn(n)ttttttnttnt	gaatgttcta(g)ttttcttgt
42459	tnnnnttttt(n)ttnttggnnt	tctagttttt(c)ttgttggcgt
42462	nnttttntt(n)ttggnntntn	agttttttctt(g)ttggcgtctg
42467	ttnttnttgg(n)ntntnnnnng	ttcttgttgg(c)gtctgctgtg
42468	tnttnttggn(n)tntnnnnngg	tcttgttggc(g)tctgctgtgg
42470	ttnttggnnt(n)tnnnnnggaa	ttgttggcgt(c)tgctgtggaa
42472	nttggnntnt(n)nnnnggaaaa	gttggcgtct(g)ctgtggaaaa
42473	ttggnntntn(n)nnnggaaaat	ttggcgtctg(c)tgtggaaaat
42474	tggnntntnn(n)nnggaaaatt	tggcgtctgc(t)gtggaaaatt
42475	ggnntntnnn(n)nggaaaattt	ggcgtctgct(g)tggaaaattt
42476	gnntntnnnn(n)ggaaaattta	gcgtctgctg(t)ggaaaattta
42489	aaaatttaaa(n)aangggnngg	aaaatttaaa(t)aatgggcagg
42492	atttaaanaa(n)gggnngggaa	atttaaataa(t)gggcagggaa
42496	aaanaanggg(n)ngggaaaata	aaataatggg(c)agggaaaata
42497	aanaangggn(n)gggaaaataa	aataatgggc(a)gggaaaataa
44079	cagagcaatc(n)atcacttgtn	cagagcaatc(a)atcacttgtg
44089	natcacttgt(n)aaaannntnt	aatcacttgt(g)acaaaagtct
44091	tcacttgtna(a)aannntntcn	tcacttgtga(c)aaaagtctgt
44094	cttgtnaaaa(n)nntntcnnca	cttgtgacaa(a)agtctgtcca
44095	ttgtnaaaan(n)ntntcnncag	ttgtgacaaa(a)gtctgtccag
44096	tgtnaaaann(n)tntcnncagn	tgtgacaaaa(g)tctgtccaga
44098	tnaaaannnt(n)tcnncagntg	tgacaaaagt(c)tgtccagatg
44100	aaaannntnt(c)nncagntgtg	acaaaagtct(g)tccagatgtg
44101	aaannntntc(n)ncagntgtgn	caaaagtctg(t)ccagatgtgt
44102	aannntntcn(n)cagntgtgnn	aaaagtctgt(c)cagatgtgtt
44106	ntntcnncag(n)tgtgnnnana	gtctgtccag(a)tgtgttgaca
44111	nncagntgtg(n)nnanantnnt	tccagatgtg(t)tgacagacttt
44112	ncagntgtgn(n)nanantnntn	ccagatgtgt(t)gacagacttg
44113	cagntgtgnn(n)anantnntnt	cagatgtgtt(g)acagacttgt
44115	gntgtgnnna(n)antnntnttt	gatgtgttga(c)agacttgtct
44117	tgtgnnnana(n)tnntntttct	tgtgttgaca(g)acttgtcttt
44118	gtgnnnanan(t)nntntttctt	gtgttgacag(a)cttgtctttc
44119	tgnnnanant(n)ntntttcttc	tgttgacaga(c)ttgtctttct
44120	gnnnanantn(n)tntttcttcg	gttgacagac(t)tgtctttctt
44122	nnanantnnt(n)tttcttcgta	tgacagactt(g)tctttcttcg
44124	anantnntnt(n)tcttcgtaca	acagacttgt(c)ttctttcgta
44126	antnntnttt(c)ttcgtacaat	agacttgtct(t)tcttcgtaca
44128	tnntnttttct(t)cgtacaatac	acttgtcttt(c)ttcgtacaat
44129	nntnttttct(c)gtacaatacg	cttgtctttc(t)tcgtacaata
44130	ntnttttctc(g)tacaatacga	ttgtctttct(t)cgtacaatac
44131	tntttcttcg(t)acaatacgag	tgtctttctt(c)gtacaatacg
44132	nttttcttcgt(a)caatacgagt	gtctttcttc(g)tacaatacga
44133	tttcttcgta(c)aatacgagtc	tctttcttcg(t)acaatacgag
44135	tcttcgtaca(a)tacgagtctn	tttcttcgta(c)aatacgagtc
44136	cttcgtacaa(t)acgagtctnn	ttcttcgtac(a)atacgagtct
44138	tcgtacaata(c)gagtctnntn	cttcgtacaa(t)acgagtctag
44139	cgtacaatac(g)agtctnntna	ttcgtacaat(a)cgagtctagt
44140	gtacaatacg(a)gtctnntnan	tcgtacaata(c)gagtctagtt
44142	acaatacgag(t)ctnntnannn	gtacaatacg(a)gtctagttaa
44143	caatacgagt(c)tnntnannnn	tacaatacga(g)tctagttaat
44145	atacgagtct(n)ntnannnnga	caatacgagt(c)tagttaatga
44146	tacgagtctn(n)tnannnngan	aatacgagtc(t)agttaatgag

Query Match 36.7%; Score 733.4; DB 8; Length 166357;
Best Local Similarity 83.7%; Pred. No. 3.7e-184;
Matches 935; Conservative 0; Mismatches 126; Indels 56; Gaps 7;

```
Qy        1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
            |||||||||||||||||| ||||  |||||||||||||||| ||||||||||||||||||
Db   147604 AAAGGCTTCTGAAATCAAACAACACCTTTCAAACTCTTATCCCAACCTCTGGAGTTGGGA 147545

Qy       61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
            |||||||||| |||||||||||||||||  ||| ||||||| ||||| || |||| |||||
Db   147544 GACATGGCTTTTCCCCTTTCTAGGTCCCATGAGAGCCATCGTGCTATTACTCGCCTTTGG 147485

Qy      121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
            ||| | |||||||||| || || |||||||||||||||||| ||||||||||||||||||||
Db   147484 GCCTTCTATTTTTAATCTTTTTGTCAAATTTGTTTCCTCCAGGATCGAGGCCATCAAGCT 147425

Qy      181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
            ||||||| |||||||||| ||||||| ||| |||||||||||||||||| |  ||||||
Db   147424 ACAGATGGTCTTACAAATGGAACCCCACATGTGCTCAACTAACAACTTCTACCAAGGACC 147365

Qy      241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
            |||||| ||||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db   147364 CCTGGATCGACCTGCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 147305

Qy      301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
            |||||||||||||||||| |||||||||||||||| |||| | || ||||||||||
Db   147304 AACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGAATTAGCCAGAGTGGTCATCGCCC 147245

Qy      361 AATCCCAACAGCAGCT--GGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGC 418
            ||| || |  |   |  |||||||||||||| |||| ||||||||||||||||||||||
Db   147244 AATTCCCAACGGCAGTTGGGGGTGTCCTGTTTAGAGGAGGGATTGAGAGGTGAAGCCAGC 147185

Qy      419 TGGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGT-- 475
            ||| ||| |||||||  |||||| |||||||||||||| |||||||||||||| | ||
Db   147184 TGGACTTCCTGGGTCGAGTGGGGGCTTGGAGAACTTTTCTGTCTAGCTAAAGGTTGGTAA 147125

Qy      476 ----------------------AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGA 512
                                   |||||||| ||||||||||||||||||||||||| |
Db   147124 ACACACCAATCAGCACTCTGTAAAAATGCACGAATCAGCACTCTGTGTCTAGCTAAAAGT 147065

Qy      513 TTGT-----------------------AAATGCACCAATCAGCACTCTGTAAAATGGA 547
            ||||                       ||||||||||||||||||||||||||||||||
Db   147064 TTGTAAATGCACTAATCAGCACTCTGTAAAAATGCACCAATCAGCACTCTGTAAAATGGA 147005

Qy      548 CCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAG 607
            |||||||||||||| ||||| || | ||||||| ||| ||||| ||||||| | |||||||
Db   147004 CCAATCAGCAGGACGTGGGTGGAGCCAAATAAAGGAATAAAAGCTGGCCAACAGAGCCAG 146945

Qy      608 CAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCA 667
            |||||||||  |||||||||||||||  |||||||||||||||||||||| | ||||||
Db   146944 CAGTGGCAATCTGCTCGGGTCCCCTTCCATGCTGTGGAAGCTTTGTTCTTTCGTTCTTCA 146885

Qy      668 CAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG-TAACAC 726
            | |||||||||||||||||| |||||| |||||||| |||||||||||||| ||||||
Db   146884 CCATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACCACCTTTATGAGCTGTTAACAC 146825

Qy      727 TCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCACGAACCCACTGGAAG 785
            |||||| || || |||||||||| ||||||||||   ||| || |||||||  | ||
Db   146824 TCACTGTGAAGGGCTGTGGCTTCACTCCTGAAGTCAGTGAGATCATGAACCCACCCGGAG 146765

Qy      786 GAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGC 845
            ||||||| ||||| |||| |  |||||||||||||||||| ||||||||||| |||||
Db   146764 GAACAAACAACTCTGGATGCACCACCTTTAAGAGCTGTAACAGTCACTGCGAAGGTCTGC 146705

Qy      846 AGCTTCACTCCTGAAGTC-AGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACA 904
            |||| |||||||||| | || ||||||| |||||||| ||||||||||||||||||||
Db   146704 GGCTTTACTCCTGAAGACAAGCGAGACCATGAACCCACCTGAAGGAAGAAACTCTGGACA 146645
```

```
Qy        905 CACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCA 964
              || ||||| ||||||||||||||||||||||||||||||||||| ||| ||||||||||||
Db     146644 CATCTGAACATCTGAAGGAACAAACTCCAGACACACCATCTTTAAGAACTGTAACACTCA 146585

Qy        965 CCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAA 1024
              || |||||||  ||||||||||||||||||||  ||||||||||||||||||||||||
Db     146584 CCACAAGGGTTCGTGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCGGAAGGAA 146525

Qy       1025 CAAATTCCAGACACAGTAGGAAATCTGTATTTTTGAT 1061
              | |||||| ||||||| |  |  | |  || ||| ||
Db     146524 CCAATTCCGGACACAGAACTAGTTATCAATATTTAAT 146488
```

Search completed: January 22, 2006, 13:38:37
Job time : 6846.68 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2006, 09:21:22 ; Search time 805.081 Seconds
(without alignments)
16564.866 Million cell updates/sec

Title: US-09-728-552A-3_COPY_59000_61000
Perfect score: 2001
Sequence: 1 aaaggcttctgaaatcagac.....cttgaagggttagagggact 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	80595	2	AAV83939	Aav83939 HC-contig
	2	1999.4	99.9	80240	2	AAV83940	Aav83940 NC-contig
	3	683.4	34.2	275449	11	ACN44194	Acn44194 Human gen
	4	683.4	34.2	325876	14	ADZ12929	Adz12929 Human can
	5	654.8	32.7	2086	12	ADQ64676	Adq64676 Novel hum
	6	619	30.9	65277	13	ABD32902	Abd32902 Human can
c	7	586	29.3	1761	3	AAC74309	Aac74309 Human sec
	8	490.6	24.5	22235	13	ABD33054	Abd33054 Human can
c	9	461.8	23.1	55763	14	AEA00519	Aea00519 Human epo
	10	453.6	22.7	249999	8	ABZ80229	Abz80229 Human tra
	11	452.6	22.6	1859	2	AAT96468	Aat96468 MSRV-1 cl
	12	452.6	22.6	1859	2	AAV43155	Aav43155 Multiple
	13	452.6	22.6	1859	9	ADB84337	Adb84337 MSRV-1 en
	14	452.6	22.6	1859	12	ADG14783	Adg14783 MSRV-1 cl
c	15	446.4	22.3	135090	11	ADN95683	Adn95683 Human BEC
c	16	445.4	22.3	4027	11	ADM02272	Adm02272 Human cDN
	17	436.8	21.8	33206	6	ABN95882	Abn95882 Gene #238
	18	436.6	21.8	28670	9	ADA03050	Ada03050 Human IQG
	19	436.6	21.8	28670	10	ADB72788	Adb72788 Human IQG
	20	436.6	21.8	28672	9	ADA66334	Ada66334 Human IQG
c	21	436.2	21.8	110000	12	ADJ25985_00	Adj25985 Human pho
c	22	435.8	21.8	152759	13	ABD33075	Abd33075 Human can
c	23	435.6	21.8	110000	6	ABX08336_00	Abx08336 Human pho
c	24	435.6	21.8	110000	12	ADN97989_00	Adn97989 Human pho
c	25	435.6	21.8	110000	12	ADO50281_00	Ado50281 Human pho
c	26	435.6	21.8	110000	14	AEB85185_00	Aeb85185 Human pho
	27	433.4	21.7	28670	11	ADL27128	Adl27128 Human gen
c	28	425.2	21.2	166181	12	ADQ20461	Adq20461 Human sof
c	29	425.2	21.2	166181	12	ADQ18633	Adq18633 Human sof
c	30	424.6	21.2	260027	11	ACN44046	Acn44046 Human gen
	31	420.4	21.0	164772	10	ADL13904	Adl13904 Osteoarth
	32	412.2	20.6	2834	3	AAZ29173	Aaz29173 Targeting
	33	412.2	20.6	6235	3	AAZ29169	Aaz29169 Human G-C
	34	412.2	20.6	6679	3	AAZ29170	Aaz29170 Partial s
	35	412.2	20.6	6679	4	AAD17443	Aad17443 Human gra
	36	412.2	20.6	6753	4	AAD17447	Aad17447 Human int
	37	401.2	20.0	2942	10	ADB62309	Adb62309 Human cDN
c	38	400.8	20.0	27666	8	ABZ74190	Abz74190 Secreted
c	39	400.8	20.0	27666	8	ADA98719	Ada98719 Human sec
c	40	400.8	20.0	27666	10	ABZ67756	Abz67756 Human sec
	41	400.8	20.0	32212	4	AAL06082	Aal06082 Human rep
	42	400.8	20.0	32212	4	ABL98647	Abl98647 Human tes
c	43	400.8	20.0	41206	8	ABZ74191	Abz74191 Secreted
c	44	400.8	20.0	41206	8	ADA98720	Ada98720 Human sec
c	45	400.8	20.0	41206	10	ABZ67757	Abz67757 Human sec
c	46	400.4	20.0	237961	6	ABQ80552	Abq80552 Human Can
	47	398.2	19.9	349938	10	ADC87621	Adc87621 Human GPC
	48	397.8	19.9	216215	10	ADF69167	Adf69167 Human MP5
	49	396.6	19.8	8455	9	AAD58741	Aad58741 Human tra
c	50	395	19.7	2612	4	AAH17215	Aah17215 Human cDN
	51	392.2	19.6	39119	8	ABZ74034	Abz74034 Secreted
	52	392.2	19.6	39119	8	ADA98641	Ada98641 Human sec
	53	392.2	19.6	39119	10	ADC20764	Adc20764 Human sec
	54	392.2	19.6	39119	10	ABZ67621	Abz67621 Human sec
c	55	389.6	19.5	2958	12	ADQ64251	Adq64251 Novel hum
c	56	388.2	19.4	3663	12	ADQ64751	Adq64751 Novel hum
c	57	388.2	19.4	4067	13	ADR08414	Adr08414 Full leng
c	58	383.8	19.2	557	12	ACH71616	Ach71616 Human gen
	59	383.8	19.2	557	12	ACH72535	Ach72535 Human gen
	60	383.8	19.2	6138	4	AAL05378	Aal05378 Human rep
	61	383.8	19.2	6138	4	ABL98244	Abl98244 Human tes
c	62	381	19.0	1000	14	AEB85683	Aeb85683 Human pho
c	63	381	19.0	110000	6	ABX08336_02	Continuation (3 of
c	64	381	19.0	110000	12	ADJ25985_02	Continuation (3 of
c	65	381	19.0	110000	12	ADN97989_02	Continuation (3 of
c	66	381	19.0	110000	12	ADO50281_02	Continuation (3 of
c	67	381	19.0	110000	14	AEB85185_02	Continuation (3 of
	68	379.4	19.0	349901	10	ADC86940	Adc86940 Human GPC
	69	376.4	18.8	1650	6	ABL49826	Abl49826 Human Ras
c	70	372.4	18.6	2289	10	ADB87406	Adb87406 Transgene
c	71	372.4	18.6	2289	12	ADJ35215	Adj35215 Human sta
c	72	372.4	18.6	2289	12	ADL96833	Adl96833 STablizin
	73	370.6	18.5	19172	4	AAL37109	Aal37109 Human mus
	74	370.6	18.5	19172	8	ABX60097	Abx60097 cDNA enco
	75	370.6	18.5	19172	12	ADJ30847	Adj30847 Human mus
c	76	370.6	18.5	169739	6	ABQ88186	Abq88186 Human ost
c	77	368.6	18.4	45121	9	ADA02744	Ada02744 Human TNF
c	78	368.6	18.4	45121	10	ADB72482	Adb72482 Human TNF
c	79	368.6	18.4	45121	10	ADC85224	Adc85224 Human Tnf
c	80	368.6	18.4	45121	12	ADM74339	Adm74339 Human car
c	81	368.6	18.4	45154	14	ADZ12594	Adz12594 Human can
c	82	368.6	18.4	95000	12	ADO56276	Ado56276 Human CAR
c	83	368.6	18.4	97400	14	ADX80724	Adx80724 Human fuc
	84	368.4	18.4	70708	12	ADQ97605	Adq97605 Human can
c	85	367.4	18.4	28818	4	AAL35901	Aal35901 Human mus
c	86	367.4	18.4	28818	8	ABX58889	Abx58889 cDNA enco
c	87	367.4	18.4	28818	12	ADJ29639	Adj29639 Human mus
	88	367.2	18.4	3305	13	ADR08314	Adr08314 Full leng
	89	367.2	18.4	65274	14	ADZ12653	Adz12653 Human can
c	90	365.6	18.3	598	12	ACH71009	Ach71009 Human gen
	91	364.4	18.2	1980	4	AAK94596	Aak94596 Human ful
	92	364.4	18.2	1980	12	ADL31502	Adl31502 Full leng

93	363.8	18.2	110000	4	AAK95240_00	Aak95240	Human neu
94	363.8	18.2	110000	4	AAK96733_00	Aak96733	Human neu
95	363.8	18.2	110000	6	ABT00010_00	Abt00010	Human neu
96	363.8	18.2	110000	6	ABT01503_00	Abt01503	Human neu
97	363.8	18.2	110000	11	ADW70291_00	Adw70291	Human neu
98	363.8	18.2	110000	12	ADH77486_00	Adh77486	Human neu
99	361.8	18.1	1043	3	AAA27494	Aaa27494	Human 5'
100	360	18.0	231222	10	ADL13693	Adl13693	Osteoarth

ALIGNMENTS

RESULT 1
AAV83939
ID AAV83939 standard; DNA; 80595 BP.
XX
AC AAV83939;
XX
DT 03-MAR-1999 (first entry)
XX
DE HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 8; Fig 6; 540pp; English.
XX
CC The present sequence represents the HC-contig derived from normal human
CC chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC produce an unusual chromosomal marker referred to as mardel(10). The
CC mardel(10) marker is mitotically stable and contains a functional
CC neocentromere at a location regarded as non-centromeric. This
CC neocentromere maps to q25.2 on chromosome 10. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in a
CC compatible cell, of replicating, acting as extra-chromosomal element and
CC segregating during cell division. The sequences can be used to construct
CC artificial chromosomes for use in gene therapy comprising a replicable,
CC segregating nucleic acid that confers a specific phenotype on cells.
CC Human artificial chromosomes can propagate in human cells and carry large
CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC they are not mutagenic. The artificial chromosomes are also useful for
CC generation of transgenic plants and animals, in production of proteins
CC and to make diagnostic reagents, e.g. for expression of cytokines,
CC receptors and growth factors, or to increase the copy number of a gene in
CC a cell. The constructs may also be used for functional and structural
CC analysis of chromosomes
XX

SQ Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

Query Match 100.0%; Score 2001; DB 2; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
Db    59000 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 59059

Qy       61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
Db    59060 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 59119

Qy      121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
Db    59120 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 59179

Qy      181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
Db    59180 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 59239

Qy      241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
Db    59240 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 59299

Qy      301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
Db    59300 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 59359

Qy      361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
Db    59360 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 59419

Qy      421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
Db    59420 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 59479

Qy      481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
Db    59480 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 59539

Qy      541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
Db    59540 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 59599

Qy      601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
Db    59600 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 59659

Qy      661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
Db    59660 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 59719

Qy      721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
Db    59720 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 59779

Qy      781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
Db    59780 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 59839

Qy      841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
Db    59840 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 59899

Qy      901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
Db    59900 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 59959

Qy      961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
Db    59960 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 60019
```

```
Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60020 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 60079

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60080 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 60139

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60140 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 60199

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60200 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 60259

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60260 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 60319

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60320 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 60379

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60380 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 60439

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60440 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 60499

Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60500 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 60559

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60560 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 60619

Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60620 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 60679

Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60680 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 60739

Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60740 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 60799

Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60800 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 60859

Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60860 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 60919

Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60920 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 60979

Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001
             |||||||||||||||||||||
Db      60980 CCTTGAAGGGTTAGAGGGACT 61000
```

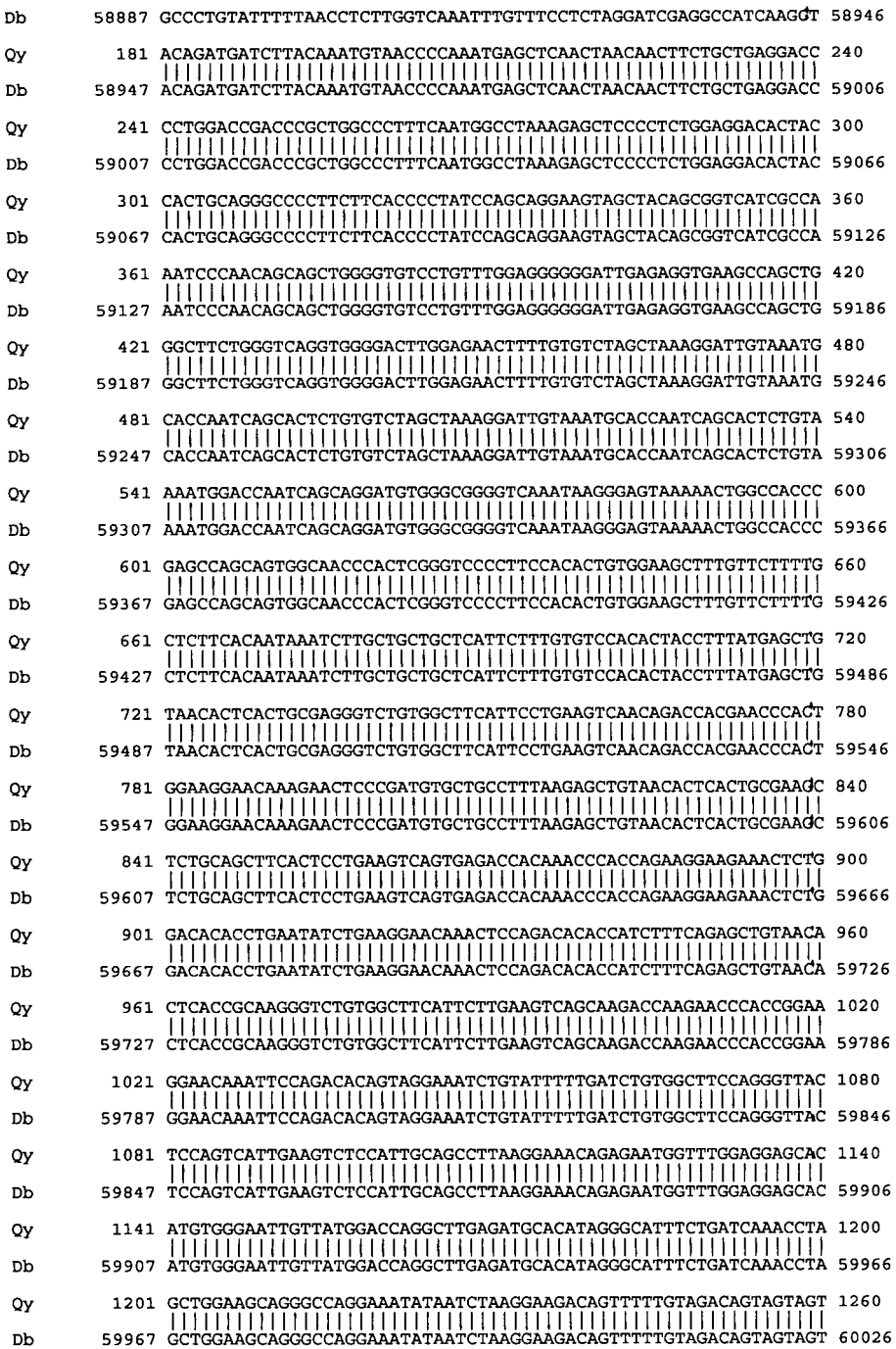
RESULT 2
AAV83940
ID AAV83940 standard; DNA; 80240 BP.
XX
AC AAV83940;
XX
DT 03-MAR-1999 (first entry)
XX
DE NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 9; Fig 16A; 540pp; English.
XX
CC The present sequence represents the NC-contig derived from a mutated
CC human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC mitotically stable and contains a functional neocentromere at a location
CC regarded as non-centromeric. This neocentromere maps to q25.2 on
CC chromosome 10. The specification describes nucleic acid sequences derived
CC from a eukaryotic chromosome, including a neocentromere or its functional
CC derivative or hybrid, that are able, in a compatible cell, of
CC replicating, acting as extra-chromosomal element and segregating during
CC cell division. The sequences can be used to construct artificial
CC chromosomes for use in gene therapy comprising a replicable, segregating
CC nucleic acid that confers a specific phenotype on cells. Human artificial
CC chromosomes can propagate in human cells and carry large amounts of DNA
CC (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC mutagenic. The artificial chromosomes are also useful for generation of
CC transgenic plants and animals, in production of proteins and to make
CC diagnostic reagents, e.g. for expression of cytokines, receptors and
CC growth factors, or to increase the copy number of a gene in a cell. The
CC constructs may also be used for functional and structural analysis of
CC chromosomes
XX
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

Query Match 99.9%; Score 1999.4; DB 2; Length 80240;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58767 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 58826

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58827 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 58886

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



RESULT 3
ACN44194
ID ACN44194 standard; DNA; 275449 BP.
XX
AC ACN44194;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG23995.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.

```
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 520; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 275449 BP; 80636 A; 53699 C; 54523 G; 86335 T; 0 U; 256 Other;

  Query Match             34.2%;  Score 683.4;  DB 11;  Length 275449;
  Best Local Similarity   80.3%;  Pred. No. 8.3e-154;
  Matches  931;  Conservative    0;  Mismatches  101;  Indels  127;  Gaps    6;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     127457 AAAAGCTTCTGATATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 127516

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
               | |||||||||||||||||||||||||  || |||||||||||||| || | || || ||
Db     127517 AATATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTATTACTTGCCTTCGG 127576

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||||||||| ||| ||||||||| |||
Db     127577 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGAATCAAGGCCATCAGGCT 127636

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||||||||||||    ||||||
Db     127637 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTATCAAGGACC 127696

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||||||| |||||||||| | ||||||||| ||||||||||||||||||||||
Db     127697 CCTGGACCGACCCACTGGCCCTTTCATTAGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 127756

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
               ||| ||||||||||||||  |||||||||||| |||||||||| || |||||||  ||
Db     127757 AACTTCAGGGCCCCTTCTTTGCCCCTATCCAGCGGGAAGTAGCTAGAGTGGTCATCACCC 127816

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || ||||| |||||| ||||||||||||||| |||||||||| ||||||||||||||||
Db     127817 AATTCCCAGCAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTAAGAGGTGAAGCCAGCT 127876

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              | |||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db     127877 GAGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAAAGGTTTGTAAAT 127936

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCA------ 533
              |||||||||||| |||||||||||||||||| ||||||||||||||||||||
Db     127937 GCACCAATCAGCGCTCTGTGTCTAGCTAAAGGTTTGTAAATGCACCAATCAGCACTTTGT 127996

Qy        534 ---------------------------------------------------------- 533
```

```
Db     127997 AAAAACACACCAATCTGCGCTCTGTGTCTAACTAAAGGTTTGTAAACGCACCGATCAGCA 128056

Qy        534 ----------------------------------------------------CTCTGTAAAAT 544
                                                                   ||||||||||
Db     128057 CTCTGTAAAAATGGACCAATCTGCGCTCTGTAAAATGGACCAATCAGTGCTCTGTAAAAT 128116

Qy        545 GGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGC 604
              |||||||||||||||| |||||| ||| |||||||||| ||||| ||||||||||||||
Db     128117 GGACCAATCAGCAGGACGTGGGCAGGGCCAAATAAGGGAATAAAAGCTGGCCACCCGAGC 128176

Qy        605 CAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCT 664
               ||||| |||||||  |||||||||||||||| |||||||||||||||||  |  ||||
Db     128177 TAGCAGCGGCAACCTGCTCGGGTCCCCTTCCACGCTGTGGAAGCTTTGTTCTCTCACTCT 128236

Qy        665 TCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAAC 724
              |||||||||||||||||||||||||  || || || ||||||||||||||||||||||||
Db     128237 TCACAATAAATCTTGCTGCTGCTCA--CTCTGGGTTCACACTACCTTTATGAGCTGTAAC 128294

Qy        725 ACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGA 783
              |||||||| || ||| || |||||||||||||||||| | |||||||||||||| ||
Db     128295 ACTCACTGTGAAGGTTTGCGGCTTCATTCCTGAAGTCAGCAAGACCACGAACCCACCGGG 128354

Qy        784 AGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCT 843
              ||||||||| ||||||    |  || |||||||| || |||||||||  | |||| |||
Db     128355 AGGAACAAATAACTCC----GCACTACCTTTAAGGGCCGTAACACTCATGGTGAAGGTCT 128410

Qy        844 GCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGAC 903
              |||||||||||||| ||||||| |||||||| |||||||  ||||| |||||  ||||
Db     128411 GCAGCTTCACTCCTAAAGTCAGCGAGACCACGAACCCACTGGAAGGACGAAACTGCGGAC 128470

Qy        904 ACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTC 963
              ||| ||||| |||    |||||||||| |||||||||||||| ||| ||||||||||
Db     128471 ACATCTGAACATCT----GAACAAACTCTGGACACACCATCTTTAAGAACTGTAACACTC 128526

Qy        964 ACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGA 1023
              || |  |||||| |||||||||||||||| ||||| |||||||||||||||| ||||||
Db     128527 ACTGTGAGGGTCCGTGGCTTCATTCTTGAAATCAGCGAGACCAAGAACCCACCAGAAGGA 128586

Qy       1024 ACAAATTCCAGACACAGTA 1042
              | |||||||||||||| ||
Db     128587 ATAAATTCCAGACACACTA 128605


RESULT 4
ADZ12929
ID   ADZ12929 standard; DNA; 325876 BP.
XX
AC   ADZ12929;
XX
DT   16-JUN-2005  (first entry)
XX
DE   Human cancer-associated genomic DNA #38.
XX
KW   Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
KW   cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2005031001-A2.
XX
PD   07-APR-2005.
XX
PF   23-SEP-2004; 2004WO-US031617.
XX
PR   23-SEP-2003; 2003US-00669920.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2005-273395/28.
XX
PT   Nucleic acid array useful for detecting cancer associated nucleic acid,
```

```
PT   comprises two or more nucleic acid probes.
XX
PS   Disclosure; SEQ ID NO 449; 198pp; English.
XX
CC   The invention relates to a nucleic acid array for detecting a cancer
CC   associated (CA) nucleic acid, comprising two or more nucleic acid probes.
CC   The invention also relates to a peptide array comprising two or more
CC   isolated polypeptides encoded by a CA nucleic acid sequence, a compound
CC   that binds to a polypeptide, an isolated antibody or its fragment which
CC   binds to a polypeptide, which is prepared by immunizing a host animal
CC   with a composition comprising the polypeptide or its antigen binding
CC   fragment and collecting cells from the host expressing antibodies against
CC   the antigen or its antigen binding fragment, a composition comprising the
CC   antibody and a carrier, a method of screening for anticancer activity, a
CC   method of detecting a CA nucleic acid, a method of diagnosing cancer, a
CC   method of treating cancer and a method of inhibiting expression of a CA
CC   nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
CC   nucleic acids. The antibody is useful for detecting the presence or
CC   absence of cancer cells in an individual which involves contacting cells
CC   from the individual with the antibody and detecting a complex of a CA
CC   protein from the cancer cells and the antibody, where the detection of
CC   the complex correlates with the presence of cancer cells in the
CC   individual. The composition is useful for inhibiting growth of cancer
CC   cells in an individual or for delivering a therapeutic agent to cancer
CC   cells in an individual. The invention is also useful for diagnosing
CC   cancer, for treating cancer and for inhibiting expression of a CA gene in
CC   a cell. This sequence represents human cancer-associated genomic DNA of
CC   the invention.
XX
SQ   Sequence 325876 BP; 95594 A; 63666 C; 65568 G; 101048 T; 0 U; 0 Other;

  Query Match             34.2%;  Score 683.4;  DB 14;  Length 325876;
  Best Local Similarity   80.3%;  Pred. No. 8.8e-154;
  Matches  931;  Conservative    0;  Mismatches  101;  Indels  127;  Gaps    6;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     127480 AAAAGCTTCTGATATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 127539

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              | |||||||||||||||||||||||  || ||||||||||||| || | || || ||
Db     127540 AATATGGCTTCTCCCCTTTCTAGGTCCCATGGCAGCCATCTTGCTATTACTTGCCTTCGG 127599

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||||||||| ||| ||||||||| |||
Db     127600 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCCTCTAGAATCAAGGCCATCAGGCT 127659

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||||||||||    ||||||
Db     127660 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTAACAACTTCTATCAAGGACC 127719

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||||||||||| |||||||||||| | |||||||||| ||||||||||||||||||||
Db     127720 CCTGGACCGACCCACTGGCCCTTTCATTAGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 127779

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              |||  |||||||||||||  |||||||||||| |||||||||| || |||||||  ||
Db     127780 AACTTCAGGGCCCCTTCTTTGCCCCTATCCAGCGGGAAGTAGCTAGAGTGGTCATCACCC 127839

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || ||||| |||||| ||||||||||||||| |||||||||| ||||||||||||||||
Db     127840 AATTCCCAGCAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTAAGAGGTGAAGCCAGCT 127899

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              | |||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db     127900 GAGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAAAGGTTTGTAAAT 127959

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCA------ 533
              |||||||||||| |||||||||||||||||| ||||||||||||||||||||
Db     127960 GCACCAATCAGCGCTCTGTGTCTAGCTAAAGGTTTGTAAATGCACCAATCAGCACTTTGT 128019

Qy        534 ------------------------------------------------------------ 533
```

```
Db     128020 AAAAACACACCAATCTGCGCTCTGTGTCTAACTAAAGGTTTGTAAACGCACCGATCAGCA 128079

Qy        534 ----------------------------------------------CTCTGTAAAAT 544
                                                           ||||||||||
Db     128080 CTCTGTAAAAATGGACCAATCTGCGCTCTGTAAAATGGACCAATCAGTGCTCTGTAAAAT 128139

Qy        545 GGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGC 604
              |||||||||||||||| |||||| ||| |||||||||| ||||| |||||||||||||
Db     128140 GGACCAATCAGCAGGACGTGGGCAGGGCCAAATAAGGGAATAAAAGCTGGCCACCCGAGC 128199

Qy        605 CAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCT 664
              ||||| ||||||| |||||||||||||||| |||||||||||||||||| | ||||
Db     128200 TAGCAGCGGCAACCTGCTCGGGTCCCCTTCCACGCTGTGGAAGCTTTGTTCTCTCACTCT 128259

Qy        665 TCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAAC 724
              |||||||||||||||||||||||||  || || || ||||||||||||||||||||||||
Db     128260 TCACAATAAATCTTGCTGCTGCTCA--CTCTGGGTTCACACTACCTTTATGAGCTGTAAC 128317

Qy        725 ACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGA 783
              |||||||| || ||| || |||||||||||||||||| | |||||||||||||| ||
Db     128318 ACTCACTGTGAAGGTTTGCGGCTTCATTCCTGAAGTCAGCAAGACCACGAACCCACCGGG 128377

Qy        784 AGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCT 843
              ||||||||| ||||||    |  || |||||||| || ||||||||||  | |||| |||
Db     128378 AGGAACAAATAACTCC----GCACTACCTTTAAGGGCCGTAACACTCATGGTGAAGGTCT 128433

Qy        844 GCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGAC 903
              |||||||||||||| ||||||| ||||||| |||||||  |||||| |||||  ||||
Db     128434 GCAGCTTCACTCCTAAAGTCAGCGAGACCACGAACCCACTGGAAGGACGAAACTGCGGAC 128493

Qy        904 ACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTC 963
              ||| ||||| |||||    |||||||||| |||||||||||||| ||| ||||||||||
Db     128494 ACATCTGAACATCT----GAACAAACTCTGGACACACCATCTTTAAGAACTGTAACACTC 128549

Qy        964 ACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGA 1023
              || |  |||||| |||||||||||||||| ||||| |||||||||||||||| ||||||
Db     128550 ACTGTGAGGGTCCGTGGCTTCATTCTTGAAATCAGCGAGACCAAGAACCCACCAGAAGGA 128609

Qy       1024 ACAAATTCCAGACACAGTA 1042
              | |||||||||||||| ||
Db     128610 ATAAATTCCAGACACACTA 128628
```

```
RESULT 5
ADQ64676
ID   ADQ64676 standard; cDNA; 2086 BP.
XX
AC   ADQ64676;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Novel human cDNA sequence #1837.
XX
KW   ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
KW   cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   EP1440981-A2.
XX
PD   28-JUL-2004.
XX
PF   21-JAN-2004; 2004EP-00001196.
XX
PR   21-JAN-2003; 2003JP-00102206.
PR   09-MAY-2003; 2003JP-00131392.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Nagai K,  Irie R;
```

```
XX
DR   WPI; 2004-535376/52.
DR   P-PSDB; ADQ66864.
XX
PT   Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 1837; 2449pp; English.
XX
CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotides and polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a nucleotide
CC   sequence of the invention.
XX
SQ   Sequence 2086 BP; 499 A; 556 C; 502 G; 529 T; 0 U; 0 Other;

  Query Match            32.7%;  Score 654.8;  DB 12;  Length 2086;
  Best Local Similarity  86.0%;  Pred. No. 1.3e-147;
  Matches  783;  Conservative    0;  Mismatches   97;  Indels   30;  Gaps    4;

Qy        158 TCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAGCTCA 217
              | |||||||||||| |||||||||||||||||||||||||||| ||||||||||||||||
Db        713 TATAGGATCGAGGCTATCAAGCTACAGATGATCTTACAAATGGAACCCCAAATGAGCTCA 772

Qy        218 ACTAACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAG 277
              ||||| ||   || |  |||||||||||||| ||||||||||||||||||||||||||||
Db        773 ACTAATAA---CTACCAAGGACCCCTGGACCAACCCGCTGGCCCTTTCAATGGCCTAAAG 829

Qy        278 AGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGA 337
              || |||||||||||||||||| ||||||||| || |||||  ||||||||||||||||
Db        830 AGTTCCCCTCTGGAGGACACTACAACTGCAGGGTCCTTTCTTTGCCCCTATCCAGCAGGA 889

Qy        338 AGTAGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGG 396
              |||||||| || |||||||  |||| |||||||||||| ||||||||| ||||  | |||
Db        890 AGTAGCTAGAGTGGTCATCACCCAATTCCCAACAGCAGTTGGGGTGTCTTGTTAAGTGGG 949

Qy        397 GGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTG 456
              | |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        950 GAGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTC 1009

Qy        457 TGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGT 516
              |||||||||| ||||||||||  |||||||||  |||||||||||||| ||| ||||||
Db       1010 TGTCTAGCTAGAGGATTGTAAACACACCAATCAGTGCTCTGTGTCTAGCTAGAGGTTTGT 1069

Qy        517 AAATGC------------------------ACCAATCAGCACTCTGTAAAATGGACCAA 551
              ||||||                        |||||||||||||||||||||||||||||
Db       1070 AAATGCACCAATCAGCACTCTGTAAAAACGGACCAATCAGCACTCTGTAAAATGGACCAA 1129

Qy        552 TCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGT 611
              |||| |||||| |||| ||| |||||||||| ||||| ||||||||| ||| |||||||
Db       1130 TCAGTAGGATGCGGGCAGGGCCAAATAAGGGAATAAAAGCTGGCCACCTGAGTCAGCAGT 1189

Qy        612 GGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAAT 671
              |||||||||||||||||||||||  ||||||||||||||||||  ||||||||||||||
Db       1190 GGCAACCCACTCGGGTCCCCTTCCATGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAAT 1249

Qy        672 AAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACT 731
              |||||||||||||||| |||||| |||||||| |||||||||||||| ||||||||||||
Db       1250 AAATCTTGCTGCTGCTCACTCTTTGGGTCCACACCACCTTTATGAGCTGCAACACTCACT 1309

Qy        732 GCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACA 790
              |||| |||||   |||||| |||||||||| | |||||||||||||  || |||||||
Db       1310 GCGAAGGTCTTCAGCTTCACTCCTGAAGTCAGCGAGACCACGAACCCATGGGGAGGAACA 1369

Qy        791 AAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTT 850
              |   ||| | ||  |||  |||||||||||||||||||||||||||||||| ||||  ||||
Db       1370 ATCGACTTCAGACATGCCACCTTTAAGAGCTGTAACACTCACTGCGAAGGTCTGTGGCTT 1429

Qy        851 CACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTG 910
              ||||||||||||||  ||||||| |||||||  |||||||||| |  ||||||| |||
Db       1430 CACTCCTGAAGTCAGCAAGACCACGAACCCACTGGAAGGAAGAAATTTCGGACACATCTG 1489

Qy        911 AATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAA 970
              || ||||||| |||||||||  ||||| |||||||| ||| |||||||||||| | |
Db       1490 AACATCTGAATGAACAAACTCTGGACACGCCATCTTTAAGAACTGTAACACTCACTGTGA 1549

Qy        971 GGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATT 1030
              ||||   ||| ||||||||||||||||||||||||||||||||| |||||||| ||||
Db       1550 GGGTTCATGGGTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCAGAAGGAACCAATT 1609

Qy       1031 CCAGACACAG 1040
              || |||||||
Db       1610 CCGGACACAG 1619


RESULT 6
ABD32902
ID   ABD32902 standard; DNA; 65277 BP.
XX
AC   ABD32902;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD18-038.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 602; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
```

```
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 65277 BP; 19651 A; 11706 C; 12664 G; 21256 T; 0 U; 0 Other;

  Query Match             30.9%;  Score 619;  DB 13;  Length 65277;
  Best Local Similarity   85.0%;  Pred. No. 1.8e-138;
  Matches  771;  Conservative    0;  Mismatches  110;  Indels   26;  Gaps    6;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||| | ||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       4321 AAAGGCTTTTTAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGAAGTTGGGC 4380

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              || |||||||||| ||||||||||||||  |||||||||||||| || || | || ||
Db       4381 AACGTGGCTTCTCCCATTTCTAGGTCCTGTTTCAGCCATCTTGCTATTACTCACCTTCGG 4440

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||  ||| |||||||||||| | |||||||||||||||| |||||| | |||||||||
Db       4441 GCTGTGTGTTTTTAACCTCCTTGTCAAATTTGTTTCCTCCAGGATCCATGCCATCAAGAC 4500

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              | ||| | |||||||||||| ||||| |||||||||||||||||||||||| |||||||||
Db       4501 ATAGA-GGTCTTACAAATGGAACCCTAAATGAGCTCAACTAACAACTTCTACTGAGGACC 4559

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||| | || | ||||||||||    ||||||| |||| ||||||||||||||||||||
Db       4560 CCTGGATCAACTCACTGGCCCTTTGGCTGGCCTAGAGAGTTCCCCTCTGGAGGACACTAC 4619

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||||||||||||||  || |||||||||||||||||||| || |||||| ||
Db       4620 CACTGCAGGGCCCCTTCTTTCCCTCTATCCAGCAGGAAGTAGCTAAAGTGGTCATAGCAC 4679

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              ||| || ||||||| | ||||||||| ||| |||||||||| || || ||||||||||||
Db       4680 AATTCCCACAGCAGTTAGGGTGTCCTATTTAGAGGGGGGATAGACAGATGAAGCCAGCTG 4739

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              || |||||||||||||||||||||||||||||||| || |||||||||| ||||||||||||
Db       4740 GGTTTCTGGGTCAGGTGGGGACTTGGAGAACTCTTCTGTCTAGCTAGAGGATTGTAAATG 4799

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              ||||| |||||| |||||                    ||||| ||||||||||||||||||
Db       4800 CACCAGTCAGCATTCTGTAA----------------AAATGGACCAATCAGCACTCTGTA 4843

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||||||||||    |||| |||| |||||||||| ||||| |||||||||
Db       4844 AAATGGACCAATCAGCAGGGCATGGGTGGGGCCAAATAAGGGAATAAAAGCTGGCCACCA 4903

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              |||||||||   ||||||  || ||||||||||||   |||||||||||| ||||||
Db       4904 GAGCCAGCAGCAGCAACCTGCTTGGGTCCCCTTCCATGCTGTGGAAGCTTTATTCTTTCG 4963

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||
Db       4964 TTCTTCATAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACTACCTTTATGAGCTG 5023

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCAC 779
              |||    |||  | |||||||||||| |||||||| ||||||| | ||||||||||||||||
```

```
Db       5024 TAA----CACCACAAGGGTCTGTGGTTTCATTCCAGAAGTCAGCGAGACCACGAACCCAC 5079

Qy        780 TGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAG 839
              |||||||||   |||||  ||| |||  |||||||| ||||||||||||||||||| ||
Db       5080 CAGAAGGAACA---AACTCTGGATATGCCACCTTTAAGGGCTGTAACACTCACTGCG-AG 5135

Qy        840 CTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCT 899
              |||||  |||||| || ||||||||| |||||||| || |||||||||||| |||  |
Db       5136 GTCTGCGGCTTCATTCTTGAAGTCAGCGAGACCACGAATCCACCAGAAGGAATAAATCCC 5195

Qy        900 GGACACA 906
              |||||||
Db       5196 AGACACA 5202


RESULT 7
AAC74309/c
ID   AAC74309 standard; cDNA; 1761 BP.
XX
AC   AAC74309;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Human secreted protein gene 30 SEQ ID NO:40.
XX
KW   Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; vulnerary; gene therapy; angiogenesis;
KW   autoimmune disease; hyperproliferative disorder; infection; skin aging;
KW   wound healing; cardiovascular disorder; cerebrovascular disorder;
KW   nervous system disorder; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200057903-A2.
XX
PD   05-OCT-2000.
XX
PF   22-MAR-2000; 2000WO-US007525.
XX
PR   26-MAR-1999;   99US-0126595P.
PR   22-DEC-1999;   99US-0171549P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-594630/56.
DR   P-PSDB; AAB39339.
XX
PT   New nucleic acid molecules encoding 48 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; Page 338; 395pp; English.
XX
CC   The polynucleotide sequences given in AAC74280 to AAC74327 encode the
CC   human secreted proteins given in AAB39310 to AAB39357. AAB39358 to
CC   AAB39400 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Examples of activities include:
CC   antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC   neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC   and vulnerary. The polynucleotides and polypeptides can be used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They can also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases, hyperproliferative disorders, cardiovascular
CC   disorders, cerebrovascular disorders, angiogenesis, nervous system
```

```
CC   disorders, infections caused by bacteria, viruses and fungi and ocular
CC   disorders. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities, fat content, lipid, protein,
CC   carbohydrate, vitamins, minerals, cofactors and other nutritional
CC   components. AAC74271 to AAC74279 and AAB39309 represent sequences used in
CC   the exemplification of the present invention
XX
SQ   Sequence 1761 BP; 409 A; 406 C; 439 G; 507 T; 0 U; 0 Other;

  Query Match             29.3%;  Score 586;  DB 3;  Length 1761;
  Best Local Similarity   77.8%;  Pred. No. 4.9e-131;
  Matches  826;  Conservative    0;  Mismatches  115;  Indels  121;  Gaps    5;

Qy        101 TTGCTAATAGTCGCATTTGGGCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCT 160
              ||  |  |  | || ||||| |||||||||||||| ||| | ||||||||||||||||||
Db       1761 TTTTTTTTTTTTGCCTTTGGACCCTGTATTTTTAATCTCCTTGTCAAATTTGTTTCCTCT 1702

Qy        161 AGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACT 220
              || || |||||||||||||||||||||| |||||||||| |||||||||||||||||||||
Db       1701 AGAATTGAGGCCATCAAGCTACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACT 1642

Qy        221 AACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGC 280
              |||||||||| |||||||||||  |||| ||| |||||||||||| ||||||||||||
Db       1641 AACAACTTCTACTGAGGACCCC-AGACCAACCTGCTGGCCCTTTCACTGGCCTAAAGAGT 1583

Qy        281 TCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGT 340
              |||||||||||||||||||| |||||||||||||||||  |||||||||||||||||| |
Db       1582 TCCCCTCTGGAGGACACTACAACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGAAAT 1523

Qy        341 AGCTACAGCGGTCATC-GCCAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGG 399
              ||||| |||||||||| ||||| |||||||| ||  || |||||||||||| || || ||
Db       1522 AGCTAGAGCGGTCATCGGCCAATTCCCAACAACACTTGTGGTGTCCTGTTTAGAAGGAGG 1463

Qy        400 ATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGT 459
              ||||||| |||||||||||||| |||||||||  |||||||||||||||||||||||| ||
Db       1462 ATTGAGAAGTGAAGCCAGCTGGACTTCTGGGTTGGGTGGGGACTTGGAGAACTTTTCTGC 1403

Qy        460 CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAA 519
              |||||||||||||||||||||||||||| ||| |||||||||||| ||||| |||||||
Db       1402 CTAGCTAAAGGATTGTAAATGCACCAATAAGCGCTCTGTGTCTAGCCAAAGGTTTGTAAA 1343

Qy        520 TGCACCAATCAGCACT---------------------------------------------- 535
              ||||||||||||||||
Db       1342 TGCACCAATCAGCACTCTGTAAAAATGCACCAATTAGCACTCTGTGTCTAACTAAAGGTT 1283

Qy        536 ------------------------------------------------------------ 535

Db       1282 TGTAAACGCACCAATCAGCACTCTGTAAAAACGGACCAATCAGCACTCTGTAAAATGGGG 1223

Qy        536 -----------CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAG 584
                        |||||||||||||||||||  |||||| |||||||| ||||||||||
Db       1222 CAATCAGCACCCTGTAAAATGGACCAATCACGAGGATGAGGGCGGGGCCAAATAAGGGAA 1163

Qy        585 TAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGG 644
              ||||| |||||||||||||||||  | |||||||  ||| |||||||||||| ||||||
Db       1162 TAAAAGCTGGCCACCCGAGCCAGTGGCGGCAACCTTCTCAGGTCCCCTTCCACGCTGTGG 1103

Qy        645 AAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACA 704
              ||||||| |||| |  |||||||||||||||||||||||||||| |||||| |||||||
Db       1102 AAGCTTTTTTCTATCACTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACA 1043

Qy        705 CTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-A 763
              |||||||||||||||||||||||||||| | ||||||  |||||| |||||||| ||
Db       1042 CTACCTTTATGAGCTGTAACACTCACTGCAAAGGTCTGCAGCTTCACTCCTGAAGCCAGT 983

Qy        764 CAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGT 823
              ||||||| ||||||| || |||||||||| |||||  || |  |  ||||||||||||||
Db        982 GAGACCACAAACCCACCGGGAGGAACAAACAACTCTGGACGCACCACCTTTAAGAGCTGT 923

Qy        824 AACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACC 883
              ||  ||||||| |||| |||| |||||||||||||||||||| |||||||| | ||||||
Db        922 AATGCTCACTGTGAAGGTCTGTGGCTTCACTCCTGAAGTCAGCGAGACCACGAGCCCACC 863

Qy        884 AGAAGGAAGAAACT---CTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACAC 940
              |||||||||||||    | |||||| ||||| |||||||   |||||||| |||||||
Db        862 AGAAGGAAGAAACTCCTCCAGACACATCTGAACATCTGAAAACACAAACTCCGGACACAC 803

Qy        941 CATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCA 1000
              ||||||| ||| ||| ||||||||| || ||||||  ||||||||||||||||||||||
Db        802 CATCTTTAAGAACTGCAACACTCACTGCGAGGGTCCATGGCTTCATTCTTGAAGTCAGCG 743

Qy       1001 AGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTA 1042
              |||||||||||||  ||||  |||| ||||||||||| | ||
Db        742 AGACCAAGAACCCCTCGGAGAGAACCAATTCCAGACATAATA 701


RESULT 8
ABD33054
ID   ABD33054 standard; DNA; 22235 BP.
XX
AC   ABD33054;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD23-007.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 858; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
```

```
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 22235 BP; 6027 A; 5133 C; 5249 G; 5826 T; 0 U; 0 Other;

  Query Match             24.5%;  Score 490.6;  DB 13;  Length 22235;
  Best Local Similarity   66.4%;  Pred. No. 1.2e-107;
  Matches  868;  Conservative    0;  Mismatches  379;  Indels   60;  Gaps    9;

Qy        323 CCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCAAATCCCAACAGCAGCTGGGGTG 382
              | | |  | |||    ||| |||    |      | |  ||  || | |||||  |  | ||
Db      12372 CACCAATCGGCACTCTGTATCTAGCTCAAGGTTTGTAAACACCAATCAGCACCCTGTGTC 12431

Qy        383 TCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGAC 442
              |       ||   | |  |   |  | |  |       ||      || |||||||||
Db      12432 TAGCTCAGGGTTTGTGACTGCACCAATCGACACTCTGCATCTAGCTACTCTGGTGGGGAC 12491

Qy        443 TTGGAGAACTTTTGTG-------TCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTC 495
              |||||||||| ||||||       | || |||    ||        | ||         ||
Db      12492 TTGGAGAACCTTTGTGTGGACACTGTATCTAGCTAATCTGGTGGGGACGTGGAGAACCTT 12551

Qy        496 TGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAA------------ 543
              |||||||||||  | |||| ||||| ||||||||||   | |||| |||
Db      12552 TGTGTCTAGCTGAGGGATGGTAAACGCACCAATCAGTGCCCTGTCAAAACAGACCACTGG 12611

Qy        544 --TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCG 601
                | ||||||||||||||||||| |||| || ||||| || ||||| | |||  ||||
Db      12612 GCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCG 12671

Qy        602 AGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGC 661
              |||||||||||||||||  ||||||||| |||||||||||||||||||||||||| ||| ||
Db      12672 AGCCAGCAGTGGCAACCGTCTCGGGTCCGCTTCCACACTGTGGAAGCTTTGTTGTTTCGC 12731

Qy        662 TCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGT 721
              ||||  ||||||||| |||||||||||  |||||| |||||||| ||||||||||||||
Db      12732 TCTTTGCAATAAATCCTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGT 12791

Qy        722 AACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACT 780
              |||||||  |||| |||||  ||||||  |||||||| || | |||||||||  |||||
Db      12792 GACACTCATGGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACCACGAGCCCACC 12851

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
              | |||||| || |||||| || |  | |||||| ||||||||||||||||||||||||
Db      12852 AGGAGGAACGAACAACTCCAGACGCACCGCCTTT-AGAGCTGTAACACTCACTGCGAAGG 12910

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
              |||||||||||||||||  || ||| |||| ||| |||||| ||||| ||||||||||
Db      12911 TCTGCAGCTTCACTCCTG-AGCCAGCGAGAGCACCAACCCAACAGAAAGAAGAAACTCCA 12969

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
              ||||| | ||| ||| ||||||| |||||||||| | ||| | || ||||||||||||
Db      12970 AACACATCCGAACATCAGAAGGAAAAAACTCCAGACGCGCCA-CCTTAAGAGCTGTAACA 13028

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
              |||||||| ||||||    |||||||||||||||||||  | ||||||||||||||
Db      13029 CTCACCGCGAGGGTCCACGGCTTCATTCTTGAAGTCAGTGACACCAAGAACCCACCAATT 13088

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
              |      |   |   || ||  | |  | ||  |       |      | ||  |   ||

Db      13089 GCGGACACACCACCACCCATAATGCCACCTTGAAACCACTTTGTCTCCGTCAGGCCCTAA 13148

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
              |    |     |||       |       |   |  || |||||      |     |  ||
Db      13149 GCTTCTTGGACAAGCGGAGAGAAGTTCAATGTGGGAACAGCAGGGAGGGCCAAACAG--- 13205

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
                                        ||  | | ||  | ||      | | | |     |
Db      13206 ------------------------TTTTGTTTCATTTTGGTTTACTTTTAACCCTTATT 13240

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
              |      |   | || || || ||    |  |||| | |  ||     |  || |||
Db      13241 ACAATCCTTATAACTAGTAAGTAGAAAAAGAATAAGAAAATGGTT------AACAGCAGT 13294

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
              ||  |  | | ||      ||          |  ||  | ||||| |   ||| |
Db      13295 GGTTTGACATAATAGGAAAAGCACTGTGCTTTAAAAAAAAAAAAAATTTCGGCCGGCCAT 13354

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
              | |||||||  |||||||||||||| |||||||||| ||| | |  ||||| ||||||
Db      13355 GGTGGCTCACACCTGTAATCCCAGCAATTTGGGAGGCCGAGGTGGGCAGATCATGAGGTC 13414

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
              ||| |||||||||||||| | |||||||  ||||||  || | ||||||||||||||||
Db      13415 AGGAGTTCGAGACCAGCCTGACTAACATGGCAAAACCCTGTTTTTACTAAAAATACAAAA 13474

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGGCAGGG 1499
              ||||||| ||  || |||||  | ||||||| ||| |||||||| ||||||||||||||
Db      13475 ATTAGCCAGGCATGCTGGCACATACCTGTAATCTCAGCTACTCAGTAGGCTGAGGCAGGA 13534

Qy       1500 GAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCC 1559
              ||||| |||||||   |||||||||||||||||||||||||| |||| |||| |||||||
Db      13535 GAATCACTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCAAGGTCACGCCACTGCACTCC 13594

Qy       1560 ATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAA 1606
              | |||||| |||||| |||||||||||||||||||| ||| ||| |
Db      13595 AGCCTGGGCAACAGAGTGAGACTCTGTCTCAAAAAAATAAATAAATA 13641


RESULT 9
AEA00519/c
ID   AEA00519 standard; DNA; 55763 BP.
XX
AC   AEA00519;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Human epoxide hydrolase (EPHX2) gene SeqID1.
XX
KW   SNP detection; haplotype mapping; nootropic; neuroprotective;
KW   Alzheimers disease; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(11006,T)
FT                   /*tag=  a
FT                   /label= Single_nucleotide_polymorphism
FT   variation       replace(21845,G)
FT                   /*tag=  b
FT                   /label= Single_nucleotide_polymorphism
FT   variation       replace(26376,A)
FT                   /*tag=  c
FT                   /label= Single_nucleotide_polymorphism
FT   variation       replace(26434,C)
FT                   /*tag=  d
FT                   /label= Single_nucleotide_polymorphism
FT   variation       replace(46768,T)
FT                   /*tag=  e
FT                   /label= Single_nucleotide_polymorphism
FT   variation       replace(54475,C)
FT                   /*tag=  f
FT                   /label= Single_nucleotide_polymorphism
```

```
XX
PN   WO2005042706-A2.
XX
PD   12-MAY-2005.
XX
PF   26-OCT-2004; 2004WO-US035235.
XX
PR   28-OCT-2003; 2003US-0515378P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Aerssens J,  Athanasious M,  Brain C,  Cohen N,  Dain B,  Denton RR;
PI   Judson RS,  Ozdemir V,  Reed CR;
XX
DR   WPI; 2005-366600/37.
XX
PT   Determining a response marker I or II by determining zero copies or at
PT   least one copy of EPHX2 haplotypes having associations with response to
PT   galantamine, useful for treating patients with Alzheimer's disease.
XX
PS   Claim 1; SEQ ID NO 1; 142pp; English.
XX
CC   This invention relates to a novel method of determining whether an
CC   individual has a response marker I or II. The method comprises
CC   determining whether the individual has zero copies or at least one copy
CC   of any of haplotypes 1-26 from 13 EPHX2 haplotypes having associations
CC   with response to galantamine and their corresponding polymorphic sites as
CC   given in the specification as Table 1. The invention may be useful for
CC   the development of compounds with a nootropic or neuroprotective activity
CC   acting as epoxide-hydrolase inhibitors. The methods and compositions of
CC   the present invention are useful for treating patients having one of the
CC   EPHX2 haplotypes, predicting the response of an individual to galantamine
CC   and for treating Alzheimers disease patients based upon their haplotype
CC   profile. The present sequence is that of the human epoxide hydrolase
CC   (EPHX2) gene of which haplotypes associated with galantamine response
CC   were identified for use in the method of the invention.
XX
SQ   Sequence 55763 BP; 13651 A; 13167 C; 13346 G; 15599 T; 0 U; 0 Other;

  Query Match             23.1%;  Score 461.8;  DB 14;  Length 55763;
  Best Local Similarity   78.2%;  Pred. No. 1.5e-100;
  Matches  639;  Conservative    0;  Mismatches  152;  Indels   26;  Gaps    6;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||| |||||||  ||||||||| ||||||||||||||||||||||||
Db       3307 AAAGGCTTCTGAAACCAGACAACACCTTTCAAATTCTTATACCAACCTCTGGAGTTGGGT 3248

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||| ||||||||||||||||||||||| |||||||||||||||||| || || | ||  |
Db       3247 GACTTGGCTTCTCCCCTTTCTAGGTCCCGTGACAGCCATCTTGCTATTACTCACCTTCAG 3188

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
               |||||||||||||||| |  |||||||||||||||| |||||| ||||||||||||||
Db       3187 TCCCTGTATTTTTAACCTCCTCCTCAAATTTGTTTCCTCCAGGATCAAGGCCATCAAGCT 3128

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||| |||| ||||||||| | ||||||
Db       3127 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTTAACTCACAACTTCTACCGAGGACT 3068

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              || ||| ||||| |||||||||| | ||||||| |||| ||||||||||||||||||||
Db       3067 GCTAGACTGACCCACTGGCCCTTTGACTGGCCTAGAGAGTTCCCCTCTGGAGGACACTAC 3008

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATC-GCC 359
              || ||||||||||||||  |||||||||||||||||||||| || |||||||  ||
Db       3007 AACGGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGTGGTCATCACCC 2948

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              |  |||||  |||||  |||||||||||||  |||||  |||||||||||||  |||||||
Db       2947 AGTTCCCAGTAGCAGGTGGGGTGTCCTGTTATGAGGGAGGATTGAGAGGTGAGGCCAGCT 2888

Qy        420 GGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGT-AA 477
              || ||| || ||||  |||||||||||| |||||||  |||||  | | || ||||| ||
Db       2887 GGACTTCCTAGGTCGAGTGGGGACTTGGGGAACTTTCCTGTCTTACAAGAGAATTGTAAA 2828

Qy        478 ATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCT 537
              |||||||||||||| ||||||     ||  ||  |      | |  |  | |   || ||
Db       2827 ATGCACCAATCAGCGCTCTGTAAAACGCACAATCAGCAGCATTCTAAAAGTAGCCAATCA 2768

Qy        538 GTAAAATGGACCAATCAGCA--------------------GGATGTGGGCGGGGTCAAA 576
                | |    ||  |                        |||    ||  | |   ||
Db       2767 CGGGGAGGATTGAAAAAAGGGCACTCTGATAGGACAGAAATGGAACATGGGAGAGGACAA 2708

Qy        577 TAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA 636
              ||||||| ||||| |||||||||| |||||| || |||||||  || ||||||||||||
Db       2707 TAAGGGAATAAAAGCTGGCCACCCCAGCCAGTAGCGGCAACCTGCTGGGGTCCCCTTCCA 2648

Qy        637 CACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTG 696
              |||||||||||||||| |||| ||| ||||| ||||| ||||||    |||| ||||||
Db       2647 TGCTGTGGAAGCTTTGTGCTTTCGCTGTTCACGATAAACCTTGCTATCACTCACTCTTTG 2588

Qy        697 TGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTG 756
              |||  | | | ||||| ||| |||||||||||||||||| ||||    ||| |||||  ||
Db       2587 GGTCTGCGCCATCTTTACGAG-TGTAACACTCACTGCGAAGGTCCACAGCTCCATTCTTG 2529

Qy        757 AAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAA 792
              |||||| | |||||||||||||| |  |||||| ||
Db       2528 AAGTCAGCAAGACCACGAACCCACCGACAGGAACCAA 2492


RESULT 10
ABZ80229
ID   ABZ80229 standard; DNA; 249999 BP.
XX
AC   ABZ80229;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human tramdorin gene region genomic DNA SEQ ID NO:26.
XX
KW   Neuroprotective; nootropic; cerebroprotective; analgesic; gene therapy;
KW   central nervous system disorder; CNS disorder; multiple sclerosis;
KW   nerve injury; neuropathic pain; stroke; trauma; non-CNS disorder; tramd;
KW   tramdorin; human; chromosome 5; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003016502-A2.
XX
PD   27-FEB-2003.
XX
PF   21-AUG-2002; 2002WO-US026637.
XX
PR   21-AUG-2001; 2001US-0313907P.
PR   21-AUG-2002; 2002US-00225810.
XX
PA   (MCLA-) MCLAUGHLIN RES INST.
XX
PI   Bermingham JR;
XX
DR   WPI; 2003-278567/27.
XX
PT   New nucleic acid sequence encoding tramdorins, e.g. mouse tramd 1, mouse
PT   tramd 2, mouse tramd 3, human tramd 1, human tramd 2, human tramd 3 or
PT   rat tramd 1, useful for treating CNS, e.g. stroke, multiple sclerosis,
PT   trauma, neuropathic pain.
XX
PS   Example 6; Fig 9; 177pp; English.
XX
CC   The present invention describes an isolated nucleic acid sequence
CC   comprising a cDNA sequence encoding mouse tramdorin (tramd) 2, mouse
CC   tramd 3, human tramd 1, human tramd 2, human tramd 3 or rat tramd 1, or
CC   the genomic sequence of mouse tramd 1 or mouse tramd 3. Mouse tramd 1 is
CC   located to chromosome 11, whereas human tramd 1 is located to chromosome
CC   5q31-33. The tramd sequences have neuroprotective, nootropic, analgesic
CC   and cerebroprotective activities, and can be used in gene therapy. The
```

```
CC   nucleic acid sequences are useful for diagnosing and treating central
CC   nervous system (CNS) disorders such as multiple sclerosis, nerve injury,
CC   neuropathic pain, stroke or trauma, and non-CNS disorders. The present
CC   sequence represents the genomic sequence of the human tramdorin gene
CC   region, which is given in the exemplification of the present invention
XX
SQ   Sequence 249999 BP; 75050 A; 54012 C; 51931 G; 67805 T; 0 U; 1201 Other;

  Query Match             22.7%;  Score 453.6;  DB 8;  Length 249999;
  Best Local Similarity   89.3%;  Pred. No. 2.3e-98;
  Matches  511;  Conservative    0;  Mismatches   59;  Indels    2;  Gaps    2;

Qy        476 AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535
              ||| |||||||||||| |||||||| |||||||||| ||||||| |||||||||||||||
Db     109411 AAACGCACCAATCAGCGCTCTGTGTTTAGCTAAAGGTTTGTAAACGCACCAATCAGCACT 109470

Qy        536 CTGT-AAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGG 594
              |||| |||||||||||||||||||| ||||| ||| |||||||||| |||||| ||||
Db     109471 CTGTAAAAATGGACCAATCAGCAGGACGTGGGTAGGGCCAAATAAGGGAATAAAAGCTGG 109530

Qy        595 CCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTT 654
              ||||  |||||||||  |||||||| |||||||||||||||  |||||||||||||||||
Db     109531 CCACTAGAGCCAGCAGCAGCAACCCATTCGGGTCCCCTTCCAAGCTGTGGAAGCTTTGTT 109590

Qy        655 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTAT 714
              |||||||||||||||||||||||||||||||||||| |||||| | ||| ||||||| |||
Db     109591 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGGCCATACTACCTGTAT 109650

Qy        715 GAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACA-GACCACGA 773
              |||||||||||| || |||| | ||||||  |||||| |||||||||||| || ||||| ||
Db     109651 GAGCTGTAACATTCGCTGCAAAGGTCTGCAGCTTCACTCCTGAAGTCAGCAAGACCATGA 109710

Qy        774 ACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACT 833
              |||||||| |||||||| |||||  || | ||  |||||||||||||||||||||||||
Db     109711 ACCCACTGGGAGGAACAAACAACTCTGGACGCGCCACCTTTAAGAGCTGTAACACTCACT 109770

Qy        834 GCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGA 893
              || ||| |||||||||||||||||||||||||  ||||||| |||||||||||||||||||
Db     109771 GCAAAGGTCTGCAGCTTCACTCCTGAAGTCAGCAAGACCACGAACCCACCAGAAGGAAGA 109830

Qy        894 AACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGC 953
              ||||| ||||||| ||||| |||||||||||||||||| |||||||||||||| ||| |
Db     109831 AACTCCGGACACATCTGAACATCTGAAGGAACAAACTCCGGACACACCATCTTTAAGAAC 109890

Qy        954 TGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCC 1013
              |||||||||||||| |||| | |||||||||||||||||||||||||||||||||||||
Db     109891 TGTAACACTCACCGCGAGGGCCCGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCC 109950

Qy       1014 ACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
              ||| |||||||| |||||| |||||| ||| |
Db     109951 ACCAGAAGGAACCAATTCCGGACACACTAGAA 109982


RESULT 11
AAT96468
ID   AAT96468 standard; cDNA; 1859 BP.
XX
AC   AAT96468;
XX
DT   26-FEB-1998  (first entry)
XX
DE   MSRV-1 clone FBd3.
XX
KW   Multiple sclerosis related virus 1; env; pol; gag; RT-PCR; primer; ss;
KW   amplification; reverse transcriptase; rheumatoid arthritis; diagnosis.
XX
OS   Multiple sclerosis associated retrovirus.
XX
PN   WO9706260-A1.
XX
PD   20-FEB-1997.
XX
PF   02-AUG-1996;   96WO-FR001244.
XX
PR   03-AUG-1995;   95FR-00009643.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B;
XX
DR   WPI; 1997-154266/14.
XX
PT   New viral material and nucleotide fragments associated with multiple
PT   sclerosis and rheumatoid arthritis - also related peptide(s) and
PT   antibodies, used for diagnosis, treatment and as vaccines.
XX
PS   Claim 1; Fig 13; 188pp; French.
XX
CC   Sequences AAT96468-77 represent claimed nucleic acid sequences derived
CC   from the multiple sclerosis related virus 1 (MSRV-1), especially from the
CC   env, pol and part of the gag gene region, but exclude the sub-region
CC   represented by AAT96527. This sequence is from clone FBd3 and is
CC   especially from the env gene. The sequence was isolated by RT-PCR using
CC   primer AAT96545. The above sequences are associated with multiple
CC   sclerosis and some are also associated with rheumatoid arthritis. Nucleic
CC   acid sequences from this virus can be used in the diagnosis of multiple
CC   sclerosis and rheumatoid arthritis
XX
SQ   Sequence 1859 BP; 635 A; 384 C; 428 G; 412 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 452.6;  DB 2;  Length 1859;
  Best Local Similarity   91.6%;  Pred. No. 7.8e-99;
  Matches  490;  Conservative    0;  Mismatches   44;  Indels    1;  Gaps    1;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||| |||||||||| || ||||||||||||||||||||||
Db       1325 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAATTCCTATACCAACCTCTGGAGTTGGGC 1384

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||| |||| ||||||||| ||||  || | |||||||||||| || |||| |||||
Db       1385 AACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGG 1444

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||| ||||||||||||||||||||||||
Db       1445 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCT 1504

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||| |||||||| |||||||||
Db       1505 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACC 1564

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||| |||| |||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db       1565 CCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 1624

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              |||||||||||||| |||| |||||||||||| |||||||||||| ||| ||||| | ||
Db       1625 CACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGTCATTGCCC 1684

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||| |||||||||||||||| |||| ||| ||||||||||||||||||||||||
Db       1685 AATTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGTGAAGCCAGCT 1744

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              || |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1745 GGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 1804

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCAC 534
              ||| |||||||  |||||||||||||||||||||||||||| ||||||||||||||
Db       1805 GCAACAATCAGTGCTCTGTGTCTAGCTAAAGGATTGTAAATACACCAATCAGCAC 1859


RESULT 12
AAV43155
ID   AAV43155 standard; cDNA; 1859 BP.
XX
```

```
AC   AAV43155;
XX
DT   29-DEC-1998  (first entry)
XX
DE   Multiple sclerosis associated retrovirus-1 clone FBd3 DNA sequence.
XX
KW   Multiple sclerosis associated retrovirus-1; MSRV-1; MS; pol gene;
KW   gag gene; env gene; rheumatoid arthritis-associated virus; ss.
XX
OS   Multiple sclerosis associated retrovirus.
XX
PN   WO9823755-A1.
XX
PD   04-JUN-1998.
XX
PF   26-NOV-1997;   97WO-IB001482.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B;
XX
DR   WPI; 1998-322732/28.
XX
PT   New nucleic acid from retroviruses - useful for diagnosis, prevention and
PT   treatment of, e.g. multiple sclerosis.
XX
PS   Example 7; Fig 13; 286pp; English.
XX
CC   The present sequence represents the multiple sclerosis (MS) associated
CC   retrovirus-1 (MSRV-1) clone FBd3 DNA sequence. The invention provides
CC   complete or partial genomic sequences of the MSRV-1 pol gene, gag gene
CC   and env gene, and polypeptides encoded by these genes. The invention also
CC   provides antibodies raised against the polypeptides. The genomic
CC   sequences, polypeptides and antibodies are also claimed useful for
CC   diagnosing infection by MS and rheumatoid arthritis-associated viruses,
CC   and also for prevention and treatment of infection with these viruses
XX
SQ   Sequence 1859 BP; 635 A; 384 C; 428 G; 412 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 452.6;  DB 2;  Length 1859;
  Best Local Similarity   91.6%;  Pred. No. 7.8e-99;
  Matches  490;  Conservative    0;  Mismatches   44;  Indels    1;  Gaps    1;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||| |||||||||| || ||||||||||||||||||||||
Db       1325 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAATTCCTATACCAACCTCTGGAGTTGGGC 1384

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
               |||||| |||| ||||||||| |||| || | |||||||||||| || |||| |||||
Db       1385 AACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGG 1444

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||| | |||||||||||||| ||||||||||||||||||||||||
Db       1445 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCT 1504

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||||| |||||||||||||||||||| |||||||| |||||||||
Db       1505 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACC 1564

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||| |||| |||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db       1565 CCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 1624

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              |||||||||||||| |||| |||||||||| |||||||||| ||| ||||| | ||
Db       1625 CACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGTCATTGCCC 1684

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||| |||||||||||||||| |||| ||| ||||||||||||||||||||||||
Db       1685 AATTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGTGAAGCCAGCT 1744

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              || |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1745 GGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 1804

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCAC 534
              ||| |||||||  |||||||||||||||||||||||||||| ||||||||||||
Db       1805 GCAACAATCAGTGCTCTGTGTCTAGCTAAAGGATTGTAAATACACCAATCAGCAC 1859


RESULT 13
ADB84337
ID   ADB84337 standard; cDNA; 1859 BP.
XX
AC   ADB84337;
XX
DT   04-DEC-2003  (first entry)
XX
DE   MSRV-1 env region clone clone FBd3.
XX
KW   MSRV; ss; multiple sclerosis; rheumatoid arthritis; gag; pol;
KW   reverse transcriptase; ribonuclease H.
XX
OS   Multiple sclerosis associated retrovirus.
XX
PN   US2003039664-A1.
XX
PD   27-FEB-2003.
XX
PF   26-NOV-1997;   97US-00979847.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (PERR/) PERRON H.
PA   (BESE/) BESEME F.
PA   (BEDI/) BEDIN F.
PA   (PARA/) PARANHOS-BACCALA G.
PA   (KOMU/) KOMURIAN-PRADEL F.
PA   (JOLI/) JOLIVET-REYNAUD C.
PA   (MAND/) MANDRAND B.
PA   (GARS/) GARSON J A.
PA   (TUKE/) TUKE P W.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2003-512253/48.
XX
PT   New isolated or purified nucleic acid associated with multiple sclerosis
PT   and/or rheumatoid arthritis, useful for detecting a virus associated with
PT   multiple sclerosis or rheumatoid arthritis in a biological sample.
XX
PS   Example 7; Fig 13; 193pp; English.
XX
CC   The invention relates to an isolated or purified nucleic acid from a
CC   virus associated with multiple sclerosis and/or rheumatoid arthritis,
CC   multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
CC   pol, gag or reverse transcriptase genes (or their fragments) encoding the
CC   proteins or defined peptides (including immunodominant peptides,
CC   antigenic peptides or conserved motifs).Also included are a process for
CC   detecting a virus associated with multiple sclerosis or rheumatoid
CC   arthritis in a biological sample, a nucleic acid probe for the detection
CC   of a virus associated with multiple sclerosis or rheumatoid arthritis, a
CC   primer for the amplification by polymerisation of a nucleic acid of a
CC   viral material associated with multiple sclerosis or rheumatoid
CC   arthritis, a polypeptide exhibiting an inhibitory activity on the
CC   proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC   and an antibody directed against the MSRV-1 virus obtained by
CC   immunologically reacting a human or animal body or cells with an
CC   immunogenic agent consisting of the antigenic polypeptide defined above.
CC   The nucleic acids are useful for detecting a biological sample a virus
CC   associated with multiple sclerosis or rheumatoid arthritis, or for
CC   detecting in a biological sample, the presence of or exposure to a virus
```

```
CC   associated with multiple sclerosis or rheumatoid arthritis. The present
CC   sequence is an MSRV cDNA clone. Note: The SEQ ID numbers for the
CC   sequences as displayed in the main body of the patent do not match the
CC   SEQ ID numbers in the sequence listing. Consequently those sequences
CC   mentioned in the claims may not be the sequences the authors intended to
CC   claim.
XX
SQ   Sequence 1859 BP; 635 A; 384 C; 428 G; 412 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 452.6;  DB 9;  Length 1859;
  Best Local Similarity   91.6%;  Pred. No. 7.8e-99;
  Matches  490;  Conservative    0;  Mismatches   44;  Indels    1;  Gaps    1;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||||| ||||||||| || ||||||||||||||||||||||
Db       1325 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAATTCCTATACCAACCTCTGGAGTTGGGC 1384

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||| |||| ||||||||| ||||  || | |||||||||||| || |||| |||||
Db       1385 AACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGG 1444

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||| ||||||||||||||||||||||
Db       1445 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCT 1504

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||| |||||||| |||||||||
Db       1505 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACC 1564

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||| |||| |||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db       1565 CCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 1624

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              ||||||||||||||| ||||  |||||||||||| |||||||||||| ||| ||||| | ||
Db       1625 CACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGTCATTGCCC 1684

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||| |||||||||||||| |||| ||| ||||||||||||||||||||||||||
Db       1685 AATTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGTGAAGCCAGCT 1744

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              || |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1745 GGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 1804

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCAC 534
              ||| |||||||  |||||||||||||||||||||||||||||| ||||||||||||
Db       1805 GCAACAATCAGTGCTCTGTGTCTAGCTAAAGGATTGTAAATACACCAATCAGCAC 1859


RESULT 14
ADG14783
ID   ADG14783 standard; cDNA; 1859 BP.
XX
AC   ADG14783;
XX
DT   26-FEB-2004  (first entry)
XX
DE   MSRV-1 clone FBd3.
XX
KW   ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX
OS   Multiple sclerosis associated retrovirus.
XX
PN   US2003198647-A1.
XX
PD   23-OCT-2003.
XX
PF   03-APR-2002; 2002US-00114104.
XX
PR   26-NOV-1996;   96US-00756429.
PR   26-NOV-1997;   97US-00979847.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2004-032461/03.
XX
PT   New isolated nucleic acid and their fragments having the pol gene of a
PT   retrovirus, useful for diagnosing, preventing and/or treating multiple
PT   sclerosis and/or rheumatoid arthritis.
XX
PS   Example 7; SEQ ID NO 42; 193pp; English.
XX
CC   The invention relates to an isolated nucleic acid which comprises the pol
CC   gene of a retrovirus associated with multiple sclerosis or rheumatoid
CC   arthritis. The methods and compositions of the present invention are
CC   useful for diagnosing, preventing and/or treating multiple sclerosis
CC   and/or rheumatoid arthritis. The present sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence 1859 BP; 635 A; 384 C; 428 G; 412 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 452.6;  DB 12;  Length 1859;
  Best Local Similarity   91.6%;  Pred. No. 7.8e-99;
  Matches  490;  Conservative    0;  Mismatches   44;  Indels    1;  Gaps    1;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||||| ||||||||| || ||||||||||||||||||||||
Db       1325 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAATTCCTATACCAACCTCTGGAGTTGGGC 1384

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||| |||| ||||||||| ||||  || | |||||||||||| || |||| |||||
Db       1385 AACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGG 1444

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||| ||||||||||||||||||||||
Db       1445 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCT 1504

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||| |||||||| |||||||||
Db       1505 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACC 1564

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||| |||| |||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db       1565 CCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 1624

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              ||||||||||||||| ||||  |||||||||||| |||||||||||| ||| ||||| | ||
Db       1625 CACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGTCATTGCCC 1684

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||| |||||||||||||| |||| ||| ||||||||||||||||||||||||||
Db       1685 AATTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGTGAAGCCAGCT 1744

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              || |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1745 GGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 1804

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCAC 534
              ||| |||||||  |||||||||||||||||||||||||||||| ||||||||||||
Db       1805 GCAACAATCAGTGCTCTGTGTCTAGCTAAAGGATTGTAAATACACCAATCAGCAC 1859


RESULT 15
ADN95683/c
ID   ADN95683 standard; DNA; 135090 BP.
XX
AC   ADN95683;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Human BEC/LEC-related gene sequence SeqID607.
XX
```

```
KW   growth; differentiation; blood endothelial cell; BEC;
KW   lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
KW   lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;
KW   vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
KW   inflammatory disease; cancer metastasis; lymphatic system; gene; ds;
KW   human.
XX
OS   Homo sapiens.
XX
PN   WO2003080640-A1.
XX
PD   02-OCT-2003.
XX
PF   07-MAR-2003; 2003WO-US006900.
XX
PR   07-MAR-2002; 2002US-0363019P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
PA   (LICN ) LICENTIA LTD.
XX
PI   Alitalo K,  Makinen T,  Petrova T,  Saharinen P,  Saharinen J;
XX
DR   WPI; 2003-876899/81.
XX
PS   Example 1; SEQ ID NO 607; 176pp; English.
XX
CC   This invention relates to a method of differentially modulating the
CC   growth or differentiation of blood endothelial cells (BEC) or lymphatic
CC   endothelial cells (LEC) comprises contacting endothelial cells with a
CC   composition comprising an agent that differentially modulates blood or
CC   lymphatic endothelial cells. Treating hereditary lymphoedema comprises
CC   identifying a human subject with lymphoedema and with a mutation in at
CC   least one allele of a gene encoding a LEC protein, where the mutation
CC   correlates with lymphoedema in human subjects, and with the proviso that
CC   the LEC protein is not VEGFR-3; and administering to the subject a
CC   composition comprising a lymphatic growth agent selected from VEGF-C or
CC   VEGF-D polypeptides and polynucleotides. The invention may be useful for
CC   the development of compounds with an antiangiogenic, cytostatic,
CC   vasotropic or antiinflammatory activity or for gene therapy. The method
CC   is useful in modulating the growth or differentiation of blood
CC   endothelial cells or lymphatic endothelial cells, in treating hereditary
CC   lymphoedema, in screening for an endothelial cell disorder or
CC   predisposition to the disorder or in monitoring the efficacy or toxicity
CC   of a drug on endothelial cells. The agent is useful in manufacturing a
CC   medicament for the differential modulation of blood vessel endothelial
CC   cell or lymphatic vessel endothelial cell growth or differentiation. The
CC   lymphatic growth agent may also be used in manufacturing a medicament for
CC   the treatment of hereditary lymphoedema resulting from a mutation in a
CC   LEC gene or of other diseases involving the lymphatic vessels, such as
CC   various inflammatory diseases and cancer metastasis via the lymphatic
CC   system. The present sequence is that of a human LEC/BEC differentially
CC   expressed gene which is related to the method of the invention.Note: This
CC   sequence does not appear in the specification but was obtained by the
CC   indexer using the source data given in table 14 of the specification.
XX
SQ   Sequence 135090 BP; 41900 A; 24202 C; 25184 G; 43804 T; 0 U; 0 Other;

  Query Match             22.3%;  Score 446.4;  DB 11;  Length 135090;
  Best Local Similarity   83.1%;  Pred. No. 1e-96;
  Matches  582;  Conservative    0;  Mismatches   56;  Indels   62;  Gaps    4;

Qy        400 ATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGT 459
              | ||||||||||||||| ||| |||||||||| |||||||||||||||||||||||||||
Db     117057 ACTGAGAGGTGAAGCCACCTGAGCTTCTGGGTCCGGTGGGGACTTGGAGAACTTTTGTGT 116998

Qy        460 CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAA 519
              ||||||||| |||||||||||||||||| |||||||||||||||| |||| ||||||||
Db     116997 CTAGCTAAAAGATTGTAAATGCACCAATAAGCACTCTGTGTCTAGCGAAAGTATTGTAAA 116938

Qy        520 TGCACCAATCAGCACTCTGTA------------------------------------ 540
              |||||||||||||||||||
Db     116937 TGCACCAATCAGCACTCTGTAAAAACGCACCAATCAGAACTCTGTAAAATGGAGCAATCA 116878

Qy        541 ----------AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAA 590
                      |||||||||||||||||||||||||||||||||| |||||||||| |||||
Db     116877 GCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGCCAAATAAGGGAATAAAAG 116818

Qy        591 CTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTT 650
              ||||||||| ||||||||| ||||||  ||| |||||||||||  ||||||||||||
Db     116817 CTGGCCACCCCAGCCAGCAGCGGCAACCTGCTCAGGTCCCCTTCCATGCTGTGGAAGCTT 116758

Qy        651 TGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCT 710
              | |||||||||| |||||||||||||||||||||||||| |||||| ||||||||||||
Db     116757 TCTTCTTTTGCTGTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACTACCT 116698

Qy        711 TTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACC 769
              ||||||||||||||||||||  |||||||||||||||| ||||||||| | | |||||
Db     116697 TTATGAGCTGTAACACTCACTGGAAGGGTCTGTGGCTTCACTCCTGAAGTTAGCGAGACC 116638

Qy        770 ACGAACCCACTGGAAG-------GAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTG 822
              |||||||||| || ||       ||||||| |||||| |||||||  ||||||||||||
Db     116637 ACGAACCCACCGGGAGGAGGGAAGAACAAACAACTCCAGATGTGCCACCTTTAAGAGCTG 116578

Qy        823 TAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCAC 882
              ||| ||||||||||||  ||||  |||||||||||| |||||||  ||||||||||||
Db     116577 TAATACTCACTGCGAAAGTCTGTGGCTTCACTCCTTAAGTCAGCAAGACCACAAACC--- 116521

Qy        883 CAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCA 942
              |||||||||||||||||||||| | ||| | ||||||||||||||||  |||||||||
Db     116520 --GAAGGAAGAAACTCTGGACACATCCGAACACCTGAAGGAACAAACTCTGGACACACCA 116463

Qy        943 TCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAG 1002
              ||||| |||||||||||||||||| | ||||||| |||||||||||||||||||||| ||
Db     116462 TCTTTAAGAGCTGTAACACTCACCGTAGGGGTCTGCGGCTTCATTCTTGAAGTCAGCGAG 116403

Qy       1003 ACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTA 1042
              |||||||||||||||||||||| ||||||  |||||| ||
Db     116402 ACCAAGAACCCACCGGAAGGAATAAATTCTGGACACAATA 116363
```

```
RESULT 16
ADM02272/c
ID   ADM02272 standard; cDNA; 4027 BP.
XX
AC   ADM02272;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human cDNA of the invention SEQ ID NO:957.
XX
KW   ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS   Homo sapiens.
XX
PN   EP1347046-A1.
XX
PD   24-SEP-2003.
XX
PF   12-APR-2002; 2002EP-00008400.
XX
PR   22-MAR-2002; 2002JP-00137785.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-723558/69.
DR   P-PSDB; ADM04715.
XX
PT   New polynucleotides and polypeptides are useful in gene therapy, for
PT   developing a diagnostic marker or medicines for regulating their
PT   expression and activity, or as a target of gene therapy.
XX
PS   Claim 1; SEQ ID NO 957; 305pp; English.
```

```
XX
CC   The invention relates to a novel human polynucleotide and the encoded
CC   polypeptide. A polynucleotide of the invention may have a use in gene
CC   therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC   as a primer for synthesizing the polynucleotide or as a probe for
CC   detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC   useful in gene therapy, for developing a diagnostic marker or medicines
CC   for regulating their expression and activity, or as a target of gene
CC   therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC   are useful as pharmaceutical agents. The present sequence represents a
CC   cDNA sequence of the invention.
XX
SQ   Sequence 4027 BP; 1001 A; 875 C; 982 G; 1169 T; 0 U; 0 Other;

  Query Match             22.3%;  Score 445.4;  DB 11;  Length 4027;
  Best Local Similarity   91.2%;  Pred. No. 5.5e-97;
  Matches  517;  Conservative    0;  Mismatches   46;  Indels    4;  Gaps    4;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTC-AAACTCTTATACCAACCTCTGGAGTTGGG 59
              ||||||||||||||||||||||| ||||||| |||||||||||||||||||||||||||
Db       1152 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAAACTCTTATACCAACCTCTGGAGTTGGG 1093

Qy         60 CGACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTG 119
              | |||||||||||||||||||||||||||||||| ||||||||||||| || |||| || |
Db       1092 CAACATGGCTTCTCCCCTTTCTAGGTCCTGTGGCAGCCATCTTGCTATTACTCGCCTTCG 1033

Qy        120 GGCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGC 179
              | |||| |||||||||||| | ||||||| |||||||||| ||||||||||||||||
Db       1032 GACCCT-TATTTTTAACCTCCTTGTCAAATCTGTTTCCTCTAGAATCGAGGCCATCAAGC 974

Qy        180 TACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGAC 239
              |||||||| |||||| |||| |||||||||||||| |||||||||||||| | ||||||
Db        973 TACAGATGGTCTTACGAATGGAACCCCAAATGAGCTAAACTAACAACTTCTACCGAGGAC 914

Qy        240 CCCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTA 299
              |||||||||||||||||||||||||| ||||||| |||||||||||||||| ||||||
Db        913 CCCTGGACCGACCCGCTGGCCCTTTCACTGGCCTAGAGAGCTCCCCTCTGGAAGACACTG 854

Qy        300 CCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-C 358
              | ||||||||| |||||||| || |||||||||||||||||||| ||| ||||||| |
Db        853 CAACTGCAGGGACCCTTCTTTGCCTCTATCCAGCAGGAAGTAGCTAGAGCAGTCATCGCC 794

Qy        359 CAAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGC 418
              ||| ||||| |||||| |||||||||||||||| ||||||||||||||||||||||||||
Db        793 CAATTCCCAGCAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGAGAGGTGAAGCCAGC 734

Qy        419 TGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              ||| ||||||||| |||||||||||||||||||| ||||||||||||||||||||||||
Db        733 TGGCCTTCTGGGTTAGGTGGGGACTTGGAGAACTTTTCTGTCTAGCTAAAGGATTGTAAA 674

Qy        479 TGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTG 538
                |||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db        673 CACACCAATCAGCACTCTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATCAGCACTCTG 614

Qy        539 T-AAAATGGACCAATCAGCAGGATGTG 564
              | |||| | ||||||||||    ||||
Db        613 TAAAAACGCACCAATCAGCGCTCTGTG 587


RESULT 17
ABN95882
ID   ABN95882 standard; DNA; 33206 BP.
XX
AC   ABN95882;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2380 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US030589.
XX
PR   02-OCT-2000; 2000US-0237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer, hepatocellular
PT   carcinoma or metastatic liver tumor in a patient, involves detecting the
PT   level of expression of two or more genes in a liver tissue sample.
XX
PS   Claim 1; SEQ ID NO 2380; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 33206 BP; 9901 A; 5689 C; 5955 G; 11661 T; 0 U; 0 Other;

  Query Match             21.8%;  Score 436.8;  DB 6;  Length 33206;
  Best Local Similarity   84.1%;  Pred. No. 1.3e-94;
  Matches  562;  Conservative    0;  Mismatches   77;  Indels   29;  Gaps    5;

Qy        402 TGAGAGGTGAAGCCAGCTGGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTC 460
              |||||||||||||||||||||| ||| |||||||   ||||||||||||||||||| ||||
Db        500 TGAGAGGTGAAGCCAGCTGGACTTCCTGGGTCGAATGGGGACTTGGAGAACTTTTCTGTC 559

Qy        461 TAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAAT 520
              |||||||||||||||||  |||||||||| |||||||||||||||||| ||||||||
Db        560 TAGCTAAAGGATTGTAAACACACCAATCAGCGCTCTGTGTCTAGCTAAAGGTTTGTAAAT 619

Qy        521 GCACCAAT------------------------CAGCACTCTGTAAAATGGACCAATCAG 555
              ||||||||                        |||||||||||||||||| ||||||||
Db        620 GCACCAATCAGCACTTTGTAAAAACAGACCAACCAGCACTCTGTAAAATGGGCCAATCAG 679

Qy        556 CAGGATGTGGGCGG-GGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGC 614
              | |||||||||||| |  |||||||||| ||||| ||||||||| |||||||| |  ||
Db        680 CGGGATGTGGGCGGAGACCAAATAAGGGAATAAAAGCTGGCCACCTGAGCCAGCCGCAGC 739

Qy        615 AACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAA 674
              ||||   || |||||||||||  |||||||| |||||||||| ||||||  |||||||
Db        740 AACCTGGTCTGGTCCCCTTCCAGGCTGTGGAAACTTTGTTCTTTCGCTCTTTGCAATAAA 799

Qy        675 TCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCG 734
              |||||||||||||   | |||| |||  |||||||||||||||||||||||||||| |||
Db        800 TCTTGCTGCTGCTCTCTTTTTGGGTCTGCACTACCTTTATGAGCTGTAACACTCACCGCG 859

Qy        735 AGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAG 793
              | |||||||||||||| |||||||||| | ||||| |||||||||   | ||||| |||
Db        860 AAGGTCTGTGGCTTCACTCCTGAAGTCAGCAAGACCTCGAACCCACCAGGAGGAAGAAAC 919

Qy        794 AACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCAC 853
```

```
                ||||||  ||  |   |    |||| ||||||||||||||||||| |||| |||||  |||||||
Db        920 AACTCCGGACGCACCATCTTTTAGAGCTGTAACACTCACTGTGAAGGTCTGCGGCTTCAC 979

Qy        854 TCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAAT 913
                ||||||||||||  ||||||  |||||||||||||||||||||||||||||||||||||
Db        980 TCCTGAAGTCAGCAAGACCAGGAACCCACCAGAAGGAAGAAACTCTGGACACATCTGAAC 1039

Qy        914 ATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGG 973
                ||| | |||||||||||||| |||||||||||||| | | ||||||  ||||||||||| |
Db       1040 ATCAG-AGGAACAAACTCCGGACACACCATCTTTAAAAACTGTAATGCTCACCGCAAGCG 1098

Qy        974 TCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCA 1033
                || |  |||||||||||||||||||||||||||||||||||  ||||||||| |||||||
Db       1099 TCCGCAGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCATTGGAAGGAACCAATTCCA 1158

Qy       1034 GACACAGT 1041
                ||||||| |
Db       1159 GACACACT 1166
```

RESULT 18
ADA03050
ID ADA03050 standard; DNA; 28670 BP.
XX
AC ADA03050;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human IQGAP1 carcinoma associated gene, SEQ ID NO:1568.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
PS Claim 1; SEQ ID NO 1568; 245pp; English.
XX
CC The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 28670 BP; 7464 A; 6666 C; 6289 G; 8251 T; 0 U; 0 Other;

```
  Query Match             21.8%;  Score 436.6;  DB 9;  Length 28670;
  Best Local Similarity   88.9%;  Pred. No. 1.4e-94;
  Matches  506;  Conservative    0;  Mismatches   59;  Indels    4;  Gaps    3;

Qy         476 AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535
               ||| ||||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||
Db       23250 AAACGCACCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAACGCACCAGTCAGCACT 23309

Qy         536 CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGC 595
               |||||||||||||| |||||||||||||||| |||| ||||||||| | ||||| |||||
Db       23310 CTGTAAAATGGACTAATCAGCAGGATGTGGGTGGGGCCAAATAAGGCAATAAAAGCTGGC 23369

Qy         596 CACCCGAGCCAGCAG-TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTT 654
               || |  |||||||||  |||||||||||||||| ||||||||  ||||||||||||||||
Db       23370 CAGCGAAGCCAGCAGCAGGCAACCCACTCGGGTGCCCTTCCATGCTGTGGAAGCTTTGTT 23429

Qy         655 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTAT 714
               |||| |||||||||| |||||||||||||||||||  || || |||| ||||||||| ||
Db       23430 CTTTCGCTCTTCACAGTAAATCTTGCTGCTGCTCA--CTCTGGGTCCGCACTACCTTCAT 23487

Qy         715 GAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACA-GACCACGA 773
               |||||||||||||| |||||||||||| |||| ||||||||||||| || |||||| |
Db       23488 GAGCTGTAACACTCAATGCGAGGGTCTGCGGCTGCATTCCTGAAGTCAGCAAGACCACAA 23547

Qy         774 ACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACT 833
               |||||| || ||||||||| |||||| |||| ||  |||||||||||||||||||||||||
Db       23548 ACCCACCGGGAGGAACAAACAACTCCAGATGAGCCACCTTTAAGAGCTGTAACACTCACT 23607

Qy         834 GCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGA 893
               |||||| | ||| |||||||||||||||||||||||||||| |||||||  |||||||||
Db       23608 GCGAAGGTGTGCGGCTTCACTCCTGAAGTCAGTGAGACCACGAACCCACTGGAAGGAAGA 23667

Qy         894 AACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGC 953
               ||||| ||||||| |||| |||||||||||||||||| ||||||||||||| |||||
Db       23668 AACTCCGGACACATCTGAACATCTGAAGGAACAAACTCCGGACACACCATCTTTAAGAGC 23727

Qy         954 TGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCC 1013
               |||||| |||||||  | |||| |||| |||||||||||||||||| |||||||| | ||
Db       23728 TGTAACGCTCACCGTGAAGGTCCGTGGTTTCATTCTTGAAGTCAGCGAGACCAAGCATCC 23787

Qy        1014 ACCGGAAGGAACAAATTCCAGACACAGTA 1042
               |||  |||||| |||||||||||| |||||
Db       23788 ACCAAAAGGAATAAATTCCAGACGCAGTA 23816
```

RESULT 19
ADB72788
ID ADB72788 standard; DNA; 28670 BP.
XX
AC ADB72788;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human IQGAP1 gene.
XX
KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS Homo sapiens.
XX
PN WO2003008583-A2.
XX
PD 30-JAN-2003.
XX
PF 26-DEC-2001; 2001WO-US051291.
XX

```
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 616; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 28670 BP; 7464 A; 6666 C; 6289 G; 8251 T; 0 U; 0 Other;

  Query Match             21.8%;  Score 436.6;  DB 10;  Length 28670;
  Best Local Similarity   88.9%;  Pred. No. 1.4e-94;
  Matches  506;  Conservative    0;  Mismatches   59;  Indels    4;  Gaps    3;

Qy        476 AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535
              ||| |||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||
Db      23250 AAACGCACCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAACGCACCAGTCAGCACT 23309

Qy        536 CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGC 595
              |||||||||||| |||||||||||||||| |||| ||||||||| | ||||| |||||
Db      23310 CTGTAAAATGGACTAATCAGCAGGATGTGGGTGGGGCCAAATAAGGCAATAAAAGCTGGC 23369

Qy        596 CACCCGAGCCAGCAG-TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTT 654
              || |  |||||||| |||||||||||||||| ||||||||  ||||||||||||||||
Db      23370 CAGCGAAGCCAGCAGCAGGCAACCCACTCGGGTGCCCTTCCATGCTGTGGAAGCTTTGTT 23429

Qy        655 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTAT 714
              |||| |||||||||| ||||||||||||||||||  || || |||| ||||||||| ||
Db      23430 CTTTCGCTCTTCACAGTAAATCTTGCTGCTGCTCA--CTCTGGGTCCGCACTACCTTCAT 23487

Qy        715 GAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACA-GACCACGA 773
              |||||||||||||| |||||||||||| |||| |||||||||||||| || |||||| |
Db      23488 GAGCTGTAACACTCAATGCGAGGGTCTGCGGCTGCATTCCTGAAGTCAGCAAGACCACAA 23547

Qy        774 ACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACT 833
              |||||| || ||||||||| |||||| |||| ||  |||||||||||||||||||||||
Db      23548 ACCCACCGGGAGGAACAAACAACTCCAGATGAGCCACCTTTAAGAGCTGTAACACTCACT 23607

Qy        834 GCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGA 893
              |||||| | ||| |||||||||||||||||||||||||||| |||||||  |||||||||
Db      23608 GCGAAGGTGTGCGGCTTCACTCCTGAAGTCAGTGAGACCACGAACCCACTGGAAGGAAGA 23667

Qy        894 AACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGC 953
              ||||| ||||||| ||||| |||||||||||||||||| |||||||||||||| |||||
Db      23668 AACTCCGGACACATCTGAACATCTGAAGGAACAAACTCCGGACACACCATCTTTAAGAGC 23727

Qy        954 TGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCC 1013
              |||||| |||||||  | |||| |||| |||||||||||||||||| |||||||| | ||
Db      23728 TGTAACGCTCACCGTGAAGGTCCGTGGTTTCATTCTTGAAGTCAGCGAGACCAAGCATCC 23787

Qy       1014 ACCGGAAGGAACAAATTCCAGACACAGTA 1042
              |||  |||||| |||||||||| |||||
Db      23788 ACCAAAAGGAATAAATTCCAGACGCAGTA 23816


RESULT 20
ADA66334
ID   ADA66334 standard; DNA; 28672 BP.
XX
AC   ADA66334;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human IQGAP1 gene genomic DNA sequence.
XX
KW   carcinoma-associated gene; CA gene; Rorc gene; mCG15938 gene; BAT1 gene;
KW   Iqgap1 gene; IQGAP1 gene; Zpf29 gene; hCG27579 gene; Kcnj9 gene;
KW   KCNJ9 gene; Ppp3cc gene; PPP3CC gene; mCG9110 gene; hCG27579 gene;
KW   cancer cell; lymphatic cell; breast cell; prostate cell; epithelial cell;
KW   carcinoma-associated protein; CAP; cytostatic; gene therapy; anticancer;
KW   vaccine; carcinoma; lymphoma carcinoma; lymphatic cancer; breast cancer;
KW   prostate cancer; DNA vaccine; animal model; human; ds; IQGAP1.
XX
OS   Homo sapiens.
XX
PN   WO2003053224-A2.
XX
PD   03-JUL-2003.
XX
PF   20-DEC-2002; 2002WO-US041776.
XX
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-569168/53.
XX
PT   Novel recombinant carcinoma-associated nucleic acid, useful for
PT   evaluating the effect of a candidate carcinoma drug, and for diagnosing
PT   carcinoma.
XX
PS   Claim 1; Page 82-86; 229pp; English.
XX
CC   This invention relates to a novel recombinant carcinoma-associated (CA)
CC   nucleic acid comprising a fully defined genomic, mRNA or coding sequences
CC   of mouse Rorc gene or human RORC gene, mouse mCG15938 or human gene BAT1,
CC   mouse Iqgap1 gene or human IQGAP1 gene, mouse Zpf29 gene or human
CC   hCG27579 gene, mouse Kcnj9 gene or human KCNJ9 gene, mouse Ppp3cc gene or
CC   human PPP3CC gene, mouse mCG9110 gene or human hCG27579 gene, as given in
CC   the specification. CA genes are genes which are preferably expressed in
CC   cancer cells, preferably lymphatic, breast, prostate or epithelial cells.
CC   A compound which modifies the expression of the CA genes or bind to
CC   carcinoma-associated proteins (CAP) may have cytostatic activity and the
CC   sequences of the invention may enable the use of gene therapy or a
CC   development of an anticancer vaccine. Therefore the invention may be
CC   useful for diagnosis and treatment of carcinomas, especially lymphoma
CC   carcinoma, breast cancer and prostate cancer. The CA genes may also be
CC   useful as DNA vaccines and for generating animal models of carcinomas.
CC   The present sequence is that of the human IQGAP1 gene genomic DNA
CC   sequence of the invention.
XX
SQ   Sequence 28672 BP; 7465 A; 6666 C; 6289 G; 8252 T; 0 U; 0 Other;

  Query Match             21.8%;  Score 436.6;  DB 9;  Length 28672;
  Best Local Similarity   88.9%;  Pred. No. 1.4e-94;
  Matches  506;  Conservative    0;  Mismatches   59;  Indels    4;  Gaps    3;

Qy        476 AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535
              ||| |||||||||||||||||||||||||||||||||| |||||| |||||| ||||||||
Db      23252 AAACGCACCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAACGCACCAGTCAGCACT 23311

Qy        536 CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGC 595
              |||||||||||| |||||||||||||||| |||| ||||||||| | ||||| |||||
Db      23312 CTGTAAAATGGACTAATCAGCAGGATGTGGGTGGGGCCAAATAAGGCAATAAAAGCTGGC 23371

Qy        596 CACCCGAGCCAGCAG-TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTT 654
              || |  |||||||| |||||||||||||||| |||||||| ||||||||||||||||
```

```
Db      23372 CAGCGAAGCCAGCAGCAGGCAACCCACTCGGGTGCCCTTCCATGCTGTGGAAGCTTTGTT 23431

Qy        655 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTAT 714
              |||| |||||||||| ||||||||||||||||||  || || |||| |||||||| ||
Db      23432 CTTTCGCTCTTCACAGTAAATCTTGCTGCTGCTCA--CTCTGGGTCCGCACTACCTTCAT 23489

Qy        715 GAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACA-GACCACGA 773
              |||||||||||||| |||||||||||| |||| |||||||||||||| || |||||| |
Db      23490 GAGCTGTAACACTCAATGCGAGGGTCTGCGGCTGCATTCCTGAAGTCAGCAAGACCACAA 23549

Qy        774 ACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACT 833
              |||||| || ||||||||| |||||| |||| ||  |||||||||||||||||||||||
Db      23550 ACCCACCGGGAGGAACAAACAACTCCAGATGAGCCACCTTTAAGAGCTGTAACACTCACT 23609

Qy        834 GCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGA 893
              |||||| | ||| |||||||||||||||||||||||||| |||||||  |||||||||
Db      23610 GCGAAGGTGTGCGGCTTCACTCCTGAAGTCAGTGAGACCACGAACCCACTGGAAGGAAGA 23669

Qy        894 AACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGC 953
              ||||| ||||||| ||||| |||||||||||||||||||| |||||||||||||| |||||
Db      23670 AACTCCGGACACATCTGAACATCTGAAGGAACAAACTCCGGACACACCATCTTTAAGAGC 23729

Qy        954 TGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCC 1013
              |||||| |||||||  | |||| |||| |||||||||||||||||| |||||||| | ||
Db      23730 TGTAACGCTCACCGTGAAGGTCCGTGGTTTCATTCTTGAAGTCAGCGAGACCAAGCATCC 23789

Qy       1014 ACCGGAAGGAACAAATTCCAGACACAGTA 1042
              |||  |||||| |||||||||||| |||||
Db      23790 ACCAAAAGGAATAAATTCCAGACGCAGTA 23818


RESULT 21
ADJ25985_00/c
WP Sequence split into 17 fragments  LOCUS ADJ25985 Accession Adj25985
WP     Fragment Name      Begin     End
WP     ADJ25985_00            1    110000
WP     ADJ25985_01       100001    210000
WP     ADJ25985_02       200001    310000
WP     ADJ25985_03       300001    410000
WP     ADJ25985_04       400001    510000
WP     ADJ25985_05       500001    610000
WP     ADJ25985_06       600001    710000
WP     ADJ25985_07       700001    810000
WP     ADJ25985_08       800001    910000
WP     ADJ25985_09       900001   1010000
WP     ADJ25985_10      1000001   1110000
WP     ADJ25985_11      1100001   1210000
WP     ADJ25985_12      1200001   1310000
WP     ADJ25985_13      1300001   1410000
WP     ADJ25985_14      1400001   1510000
WP     ADJ25985_15      1500001   1610000
WP     ADJ25985_16      1600001   1691139
ID   ADJ25985 standard; DNA; 1691139 BP.
XX
AC   ADJ25985;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human phosphodiesterase 4D (PDE4D) genomic DNA.
XX
KW   Human; phosphodiesterase 4D; PDE4D; gene; ds; stroke; at-risk haplotype;
KW   polymorphism; cerebroprotective; vasotropic; chromosome 5q12.
XX
OS   Homo sapiens.
XX
PN   US2004014099-A1.
XX
PD   22-JAN-2004.
XX
PF   18-APR-2003; 2003US-00419723.
XX
PR   19-MAR-2001; 2001US-00811352.
PR   04-FEB-2002; 2002US-00067514.
PR   25-SEP-2002; 2002US-00255120.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Gretarsdottir S,  Jonsdottir S,  Reynisdottir ST,  Thorleifsson G;
PI   Gulcher JR;
XX
DR   WPI; 2004-121560/12.
XX
PT   Diagnosing susceptibility to a stroke in an individual, comprises
PT   screening for an at-risk haplotype in the phosphodiesterase 4D gene.
XX
PS   Claim 5; SEQ ID NO 1; 417pp; English.
XX
CC   The invention relates to a method of diagnosing susceptibility to a
CC   stroke in an individual comprising screening for an at-risk haplotype in
CC   the phosphodiesterase 4D (PDE4D) gene that is more frequently present in
CC   an individual susceptible to stroke compared to a healthy individual,
CC   where the presence of the at-risk haplotype is indicative of a
CC   susceptibility to stroke and where the at-risk haplotype increases risk
CC   of stroke significantly. Polymorphisms in the phosphodiesterase 4D
CC   (PDE4D) gene are useful in diagnosing susceptibility to stroke in an
CC   individual, predicting the clinical course and treatment for stroke and
CC   monitoring the effectiveness of treatment. PDE4D nucleic acids,
CC   polypeptides or agents that alter the activity of the polypeptides are
CC   useful for treating or preventing stroke. This sequence represents human
CC   PDE4D genomic DNA of the invention. Note: The sequence data for this
CC   patent is also available in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 1691139 BP; 511168A; 310872C; 323134G; 545521T; 0U; 444Other;

  Query Match             21.8%;  Score 436.2;  DB 12;  Length 110000;
  Best Local Similarity   67.1%;  Pred. No. 2.7e-94;
  Matches  803;  Conservative    1;  Mismatches  289;  Indels  104;  Gaps    9;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              || |||||||| | ||||||||||||||||||||   |||||||||| ||||||||||
Db      99342 TCTGGTGGGGAGGTAGAGAACTTTTGTGTCTAGCTCGGGGATTGTAAACGCACCAATCAG 99283

Qy        491 CACTCTGT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAA 542
              ||| ||||        | | |     |  |||   ||||||| || |||| ||||
Db      99282 CACCCTGTCAAAACAGACCAATCAGCTCTCTGTAAAACAGACCAATCGGCTCTCTTTAAA 99223

Qy        543 ATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGA 602
              |||||||||||||||||| ||||| |||| || ||||| || ||||| | |||  ||| |
Db      99222 ATGGACCAATCAGCAGGACGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCAA 99163

Qy        603 GCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCT 662
              |||||||||| ||||  ||  ||| ||||| ||  |||||| |||||||||||||  ||
Db      99162 GCCAGCAGTGGGAACCTGCTGAGGTTCCCTTTCATGCTGTGGTAGCTTTGTTCTTTCACT 99103

Qy        663 CTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTA 722
              |||  |||||||||||| || |||||| |||||| |||||||| |||||||||||||||
Db      99102 CTTTGCAATAAATCTTGTTGTTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTA 99043

Qy        723 ACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTG 781
              ||||||||:  || ||||||  ||||||  |||||||  | | | ||| ||||||| | |
Db      99042 ACACTCACYATGAAGGTCTGCAGCTTCACCCCTGAAGCTAGCAAAACCGCGAACCCGCCG 98983

Qy        782 GAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCT 841
              | |||||| || |||||||||  || || | |||||||||||||||||||||||||||| |
Db      98982 GGAGGAACGAACAACTCCCGACATG-TGGCCTTAAGAGCTGTAACACTCACTGCGAAGGT 98924

Qy        842 CTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGG 901
              |||||||||| ||||| || ||| ||||||   |||||||||||||| |||||||| |
Db      98923 TTGCAGCTTCATTCCTG-AGCCAGCGAGACCGTGAACCCACCAGAAGAAAGAAACTCCGA 98865

Qy        902 ACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACAC 961
              ||||| |  || ||| |||||||||||||||||||| |   |||| ||| ||||||||
Db      98864 ACACATCCAAACATCAGAAGGAACAAACTCCAGACACGCTGGCTTTAAGAACTGTAACAT 98805

Qy        962 TCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAG 1021
```

```
Db      98804 TCACTGAGAGGGTCTGCGGCTTCATTCTTGAAATCAGTAAGACCAAGAACCCACC----- 98750

Qy       1022 GAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACT 1081

Db      98749 -----AACTCAGGACACAGTGTGA---------------------------------- 98731

Qy       1082 CCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACA 1141

Db      98730 -------------------------TTTTCACAGCCGGAAATTGACTTGGAAATCCTTA 98697

Qy       1142 TGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAG 1201

Db      98696 AGTTACTTTTTTTTTAATGCAGCTTTTGGATAGCCATAAAAATCTTATACCTTCCTTTAA 98637

Qy       1202 CTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTC 1261

Db      98636 TGTTATTTAAATTTTTGAAATAAGTTCAGTAACAAGAAAACAAGATGGACACAA------ 98583

Qy       1262 TTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCG 1321

Db      98582 -CAGTAACTTTAAAACGATTTTCTTTAGATCTTAAAAGATACGTATTTCAGCCAGGCATG 98524

Qy       1322 ATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCA 1381

Db      98523 GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCACGAGGTCA 98464

Qy       1382 GGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACA---- 1437

Db      98463 GGAGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGCCTCTAGTAAAAATACAAAAA 98404

Qy       1438 ---------AAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAG 1487

Db      98403 AAAAAAAAAAAAATTAGCCCGGAGTGGGGGCAGGCGCCTGTAGTCCTGGCTACTCGGGAG 98344

Qy       1488 GCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACA 1547

Db      98343 GCTGAGGCAGGAGAATAGCATGAACCCAGGAGGCGGAGCTTGCAATAAGCCGAGATTGTG 98284

Qy       1548 CCACAGCACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAA 1604

Db      98283 CCATTGCACTCCAGCCTGGGTGACAGTGTGAGACTCCATCTCAAAAAAAAAAAGATA 98227
```

RESULT 22
ABD33075/c
ID ABD33075 standard; DNA; 152759 BP.
XX
AC ABD33075;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-002.
XX
KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-499109/47.
XX
PT Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 10; 182pp; English.
XX
CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 152759 BP; 43847 A; 29952 C; 31446 G; 47160 T; 0 U; 354 Other;

Query Match 21.8%; Score 435.8; DB 13; Length 152759;
Best Local Similarity 66.4%; Pred. No. 3.8e-94;
Matches 872; Conservative 0; Mismatches 287; Indels 154; Gaps 11;

```
Qy       431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490

Db     92436 TCCAGTGGGGAGGTGGAGAACTTTTGTGTCTAGCTCAGGGATTGTAAATGCACCAATCAG 92377

Qy       491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCA 550

Db     92376 CACCCTGACA----------------AAACAGACCAATCAGCTCTCTGTAAAATGGACCA 92333

Qy       551 ATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAG 610

Db     92332 ATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCTGAGCCAGCAG 92273

Qy       611 TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAA 670

Db     92272 TGGCAACCTGCTGGGGTCCCCTTCCATACTGTGGAAGCTTTGTTCTCTCGCTCTTTGCAG 92213

Qy       671 TAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCAC 730

Db     92212 TAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTAACACTCAC 92153

Qy       731 TGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAAC 789

Db     92152 TGGGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACCACGAACCCACCGCAAGGAAC 92093

Qy       790 AAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCT 849

Db     92092 GAACAACTCCAGACATGCCGCC-TTAAGAGCTGCAACACTCACCGCAAAGGTCCGCAGCT 92034

Qy       850 TCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCT 909

Db     92033 TCACTCCTG-AGCCAACGAGACCACGAACCTACCAGAAGGAAGAAACTCCGAACACATCT 91975

Qy       910 GAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCA 969

Db     91974 GAACATCAGAAGGAACAAACTCCAGACACGACACCTTTAAGAACTGTAACACTTACCGCG 91915

Qy       970 AGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC-------GGAAGG 1022

Db     91914 AGGGTCCATGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCAATTCCGGATACG 91855

Qy      1023 AACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTT------------ 1070

Db     91854 CTACCATGTTTTACAGACTCGAAAATATCTTTTTTTTTTTTTTTTTTTTACCTTTAACATG 91795

Qy      1071 -------------CCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAA 1117
```

```
Db      91794 TTTGAGATCTGGACTCAGCATTCCACAATTGTTATTGGCTAAGCAGCAGTTATGGCAAAG 91735

Qy       1118 CAGAGAATGGTTTGGAGGAGCACATGTGGGAATTGTTAT---------------------- 1156

Db      91734 CTGTCATTGCCAAACGTCATAGCTGTTGACACTGATGATACTTCCAGTGGCTATGGGCAT 91675

Qy       1157 ---GGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCTGGAAGCAGGG 1213

Db      91674 TGGTGACACATGTTGAGTGTGCCATGTTAAATGTCTTTAAAAACTTATGATTTAGCACTG 91615

Qy       1214 CCAGGAAATATAA--TCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTTGCATCTG 1271

Db      91614 AGAAGAAAAGTTACTGCACAGAAAAGCACAAAAGCAGGAGCATGGCATGTATTTGATGTT 91555

Qy       1272 AGACATGTAGATTATCA---------------------------------------------- 1288

Db      91554 AGTGAAACAAATATTTATAAGACTATTAAACAAAAGCAAGTCAATGTTTTCTTGCAAAGC 91495

Qy       1289 ------------------------------AGCAATTAATTAGAAAAAATATAGCCAGG 1317

Db      91494 AACAACCAAGCCCTCTAATGAAACCCACAATAGGAAGAAACCTTTAAAAAGATAAAGCTG 91435

Qy       1318 TGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCAC--G 1375

Db      91434 GCTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGTGGGGCAGATCACCTG 91375

Qy       1376 AGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATA 1435

Db      91374 AGGTCAGGAGTTCAAGACAAGCCTGACCAACATGGTGAAACCCCATCTCTACAAAAAATA 91315

Qy       1436 CAAAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGG 1494

Db      91314 CAAAAAGTAGCCGAGCATGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGG 91255

Qy       1495 CAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGC 1554

Db      91254 CAGGAGAATCGCTTGAACCTGGGAGACAGAGGTTGCAGTGAGCCGAGATTGCACCACTGC 91195

Qy       1555 ACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAA 1607

Db      91194 ACTCCAGCCTGGGAAACAGAGCAAGACTCCGTCTCAATAAACAAAAACAAAAA 91142
```

```
RESULT 23
ABX08336_00/c
WP Sequence split into 17 fragments  LOCUS ABX08336 Accession Abx08336
WP     Fragment Name      Begin     End
WP     ABX08336_00            1    110000
WP     ABX08336_01       100001    210000
WP     ABX08336_02       200001    310000
WP     ABX08336_03       300001    410000
WP     ABX08336_04       400001    510000
WP     ABX08336_05       500001    610000
WP     ABX08336_06       600001    710000
WP     ABX08336_07       700001    810000
WP     ABX08336_08       800001    910000
WP     ABX08336_09       900001   1010000
WP     ABX08336_10      1000001   1110000
WP     ABX08336_11      1100001   1210000
WP     ABX08336_12      1200001   1310000
WP     ABX08336_13      1300001   1410000
WP     ABX08336_14      1400001   1510000
WP     ABX08336_15      1500001   1610000
WP     ABX08336_16      1600001   1691080
ID   ABX08336 standard; DNA; 1691080 BP.
XX
AC   ABX08336;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human phosphodiesterase 4D (PDE4D) gene.
XX
KW   Human; Phosphodiesterase 4D; PDE4D; stroke; cerebroprotective; gene; ds;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       732790
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       735966
FT                   /*tag= b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       736226
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       736516
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       850001
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       852776
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       853079
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       853575
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       856468
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       860845
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       870924
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1027267
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1027643
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1027757
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1028146
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1037657
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1044016
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1044045
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1254737
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1254849
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1255763
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1257206
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1258161
```

```
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1267187
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1268007
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1268553
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1272669
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1272910
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1273023
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1273220
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1273240
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1273543
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1288439
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1289730
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1290176
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1293745
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1344605
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1344864
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1345135
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1345286
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1346112
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1352976
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1354291
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1354377
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1354554
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1354675
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1355114
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1355693
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1357081
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1362985
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1363021
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1363827
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1363911
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1364061
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1364066
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1367904
FT                   /*tag= bd

  Query Match             21.8%;  Score 435.6;  DB 6;  Length 110000;
  Best Local Similarity   67.1%;  Pred. No. 3.8e-94;
  Matches  803;  Conservative    0;  Mismatches  290;  Indels  104;  Gaps    9;

Qy       431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
            || ||||||||  | ||||||||||||||||||||    |||||||||| ||||||||||
Db     99343 TCTGGTGGGGAGGTAGAGAACTTTTGTGTCTAGCTCGGGGATTGTAAACGCACCAATCAG 99284

Qy       491 CACTCTGT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAA 542
            |||  ||||         | | |       |   |||   ||||||| || |||| ||||
Db     99283 CACCCTGTCAAAACAGACCAATCAGCTCTCTGTAAAACAGACCAATCGGCTCTCTTTAAA 99224

Qy       543 ATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGA 602
            |||||||||||||||||| ||||| |||| || ||||| || ||||| | |||  ||| |
Db     99223 ATGGACCAATCAGCAGGACGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCAA 99164

Qy       603 GCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCT 662
            |||||||||| ||||  ||  ||| ||||| ||  |||||| ||||||||||||||  ||
Db     99163 GCCAGCAGTGGGAACCTGCTGAGGTTCCCTTTCATGCTGTGGTAGCTTTGTTCTTTCACT 99104

Qy       663 CTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTA 722
            |||  |||||||||| || |||||| |||||| ||||||||| ||||||||||||||||
Db     99103 CTTTGCAATAAATCTTGTTGTTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTA 99044

Qy       723 ACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTG 781
            |||||||   || ||||||  ||||||  |||||||  | | | ||| ||||||| | |
Db     99043 ACACTCACNATGAAGGTCTGCAGCTTCACCCCTGAAGCTAGCAAAACCGCGAACCCGCCG 98984

Qy       782 GAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCT 841
            | |||||| || |||||||||  || || | ||||||||||||||||||||||||||||| |
Db     98983 GGAGGAACGAACAACTCCCGACATG-TGGCCTTAAGAGCTGTAACACTCACTGCGAAGGT 98925

Qy       842 CTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGG 901
            ||||||||||  |||| || ||| ||||||   |||||||||||| ||||||||| |
Db     98924 TTGCAGCTTCATTCCTG-AGCCAGCGAGACCGTGAACCCACCAGAAGAAAGAAACTCCGA 98866

Qy       902 ACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACAC 961
            |||||  |  || ||| |||||||||||||||||||  |   |||| ||| ||||||||
Db     98865 ACACATCCAAACATCAGAAGGAACAAACTCCAGACACGCTGGCTTTAAGAACTGTAACAT 98806

Qy       962 TCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAG 1021
            |||| |  |||||||| |||||||||||||| |||| ||||||||||||||||
Db     98805 TCACTGAGAGGGTCTGCGGCTTCATTCTTGAAATCAGTAAGACCAAGAACCCACC----- 98751

Qy      1022 GAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACT 1081
```

```
Db      98750 -----AACTCAGGACACAGTGTGA---------------------------------- 98732

Qy       1082 CCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACA 1141

Db      98731 -------------------------TTTTCACAGCCGGAAATTGACTTGGAAATCCTTA 98698

Qy       1142 TGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAG 1201

Db      98697 AGTTACTTTTTTTTTAATGCAGCTTTTGGATAGCCATAAAAATCTTATACCTTCCTTTAA 98638

Qy       1202 CTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTC 1261

Db      98637 TGTTATTTAAATTTTTGAAATAAGTTCAGTAACAAGAAAACAAGATGGACACAA------ 98584

Qy       1262 TTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCG 1321

Db      98583 -CAGTAACTTTAAAACGATTTTCTTTAGATCTTAAAAGATACGTATTTCAGCCAGGCATG 98525

Qy       1322 ATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCA 1381

Db      98524 GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCACGAGGTCA 98465

Qy       1382 GGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACA---- 1437

Db      98464 GGAGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGCCTCTAGTAAAAATACAAAAA 98405

Qy       1438 ---------AAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAG 1487

Db      98404 AAAAAAAAAAAAATTAGCCCGGAGTGGGGGCAGGCGCCTGTAGTCCTGGCTACTCGGGAG 98345

Qy       1488 GCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACA 1547

Db      98344 GCTGAGGCAGGAGAATAGCATGAACCCAGGAGGCGGAGCTTGCAATAAGCCGAGATTGTG 98285

Qy       1548 CCACAGCACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAA 1604

Db      98284 CCATTGCACTCCAGCCTGGGTGACAGTGTGAGACTCCATCTCAAAAAAAAAAAGATA 98228
```

RESULT 24
ADN97989_00/c
WP Sequence split into 17 fragments LOCUS ADN97989 Accession Adn97989

	Fragment Name	Begin	End
WP	ADN97989_00	1	110000
WP	ADN97989_01	100001	210000
WP	ADN97989_02	200001	310000
WP	ADN97989_03	300001	410000
WP	ADN97989_04	400001	510000
WP	ADN97989_05	500001	610000
WP	ADN97989_06	600001	710000
WP	ADN97989_07	700001	810000
WP	ADN97989_08	800001	910000
WP	ADN97989_09	900001	1010000
WP	ADN97989_10	1000001	1110000
WP	ADN97989_11	1100001	1210000
WP	ADN97989_12	1200001	1310000
WP	ADN97989_13	1300001	1410000
WP	ADN97989_14	1400001	1510000
WP	ADN97989_15	1500001	1610000
WP	ADN97989_16	1600001	1691138

ID ADN97989 standard; DNA; 1691138 BP.
XX
AC ADN97989;
XX
DT 01-JUL-2004 (first entry)
XX
DE Human phosphodiesterase 4D genomic sequence.
XX
KW cerebroprotective; vasotropic; PDE4D modulator;
KW phosphodiesterase 4D modulator; phosphodiesterase-4D-Modulator; stroke;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004028341-A2.
XX
PD 08-APR-2004.
XX
PF 25-SEP-2003; 2003WO-US029906.
XX
PR 25-SEP-2002; 2002US-00255120.
PR 18-APR-2003; 2003US-00419723.
PR 27-AUG-2003; 2003US-00650120.
XX
PA (DECO-) DECODE GENETICS EHF.
XX
PI Gretarsdottir S, Reynisdottir ST, Thorleifsson G, Gulcher JR;
XX
DR WPI; 2004-315987/29.
XX
PT Assessing or diagnosing susceptibility to stroke, comprises determining
PT phosphodiesterase 4D isoform expression levels in an individual compared
PT to control, a difference in isoform expression indicates susceptibility
PT to stroke.
XX
PS Example 1; SEQ ID NO 105; 521pp; English.
XX
CC The invention relates to a method of assessing susceptibility to stroke
CC in an individual by determining phosphodiesterase 4D (PDE4D) isoform
CC expression levels in the individual compared to control, where a
CC difference in isoform expression indicates susceptibility to stroke. The
CC method is useful for assessing susceptibility to stroke (claimed), for
CC predicting those at higher risk of developing a stroke, for specifically
CC identifying a rate-limiting pathway involved in stroke, for predicting an
CC individual's response to a particular drug, even drugs that do not act on
CC PDE4D or its pathway. Modulators of PDE4D isoform are useful for
CC preventing the occurrence of stroke in an individual. PDE4D polypeptides
CC are useful as molecular weight markers, to raise antibodies or to elicit
CC an immune response, or as markers for cells or tissues in which the
CC corresponding polypeptide is preferentially expressed during tissue
CC differentiation or in a diseased state. This sequence corresponds to the
CC PDE4D genomic sequence.
XX
SQ Sequence 1691138 BP; 511166A; 310874C; 323132G; 545522T; 0U; 444Other;

Query Match 21.8%; Score 435.6; DB 12; Length 110000;
Best Local Similarity 67.1%; Pred. No. 3.8e-94;
Matches 803; Conservative 0; Mismatches 290; Indels 104; Gaps 9;

```
Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490

Db      99343 TCTGGTGGGGAGGTAGAGAACTTTTGTGTCTAGCTCGGGGATTGTAAACGCACCAATCAG 99284

Qy        491 CACTCTGT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAA 542

Db      99283 CACCCTGTCAAAACAGACCAATCAGCTCTCTGTAAAACAGACCAATCGGCTCTCTTTAAA 99224

Qy        543 ATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGA 602

Db      99223 ATGGACCAATCAGCAGGACGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCAA 99164

Qy        603 GCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCT 662

Db      99163 GCCAGCAGTGGGAACCTGCTGAGGTTCCCTTTCATGCTGTGGTAGCTTTGTTCTTTCACT 99104

Qy        663 CTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTA 722

Db      99103 CTTTGCAATAAATCTTGTTGTTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTA 99044

Qy        723 ACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTG 781

Db      99043 ACACTCACNATGAAGGTCTGCAGCTTCACCCCTGAAGCTAGCAAAACCGCGAACCCGCCG 98984

Qy        782 GAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCT 841
```

```
Db      98983 GGAGGAACGAACAACTCCCGACATG-TGGCCTTAAGAGCTGTAACACTCACTGCGAAGGT 98925

Qy        842 CTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGG 901

Db      98924 TTGCAGCTTCATTCCTG-AGCCAGCGAGACCGTGAACCCACCAGAAGAAAGAAACTCCGA 98866

Qy        902 ACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACAC 961

Db      98865 ACACATCCAAACATCAGAAGGAACAAACTCCAGACACGCTGGCTTTAAGAACTGTAACAT 98806

Qy        962 TCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAG 1021

Db      98805 TCACTGAGAGGGTCTGCGGCTTCATTCTTGAAATCAGTAAGACCAAGAACCCACC----- 98751

Qy       1022 GAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACT 1081

Db      98750 -----AACTCAGGACACAGTGTGA---------------------------------- 98732

Qy       1082 CCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACA 1141

Db      98731 --------------------------TTTTCACAGCCGGAAATTGACTTGGAAATCCTTA 98698

Qy       1142 TGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAG 1201

Db      98697 AGTTACTTTTTTTTTAATGCAGCTTTTGGATAGCCATAAAAATCTTATACCTTCCTTTAA 98638

Qy       1202 CTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTC 1261

Db      98637 TGTTATTTAAATTTTTGAAATAAGTTCAGTAACAAGAAAACAAGATGGACACAA------ 98584

Qy       1262 TTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCG 1321

Db      98583 -CAGTAACTTTAAAACGATTTTCTTTAGATCTTAAAAGATACGTATTTCAGCCAGGCATG 98525

Qy       1322 ATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCA 1381

Db      98524 GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCACGAGGTCA 98465

Qy       1382 GGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACA---- 1437

Db      98464 GGAGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGCCTCTAGTAAAAATACAAAAA 98405

Qy       1438 ---------AAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAG 1487

Db      98404 AAAAAAAAAAAAATTAGCCCGGAGTGGGGGCAGGCGCCTGTAGTCCTGGCTACTCGGGAG 98345

Qy       1488 GCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACA 1547

Db      98344 GCTGAGGCAGGAGAATAGCATGAACCCAGGAGGCGGAGCTTGCAATAAGCCGAGATTGTG 98285

Qy       1548 CCACAGCACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAA 1604

Db      98284 CCATTGCACTCCAGCCTGGGTGACAGTGTGAGACTCCATCTCAAAAAAAAAAAGATA 98228
```

```
RESULT 25
ADO50281_00/c
WP Sequence split into 17 fragments  LOCUS ADO50281 Accession Ado50281
WP      Fragment Name       Begin      End
WP      ADO50281_00             1    110000
WP      ADO50281_01        100001    210000
WP      ADO50281_02        200001    310000
WP      ADO50281_03        300001    410000
WP      ADO50281_04        400001    510000
WP      ADO50281_05        500001    610000
WP      ADO50281_06        600001    710000
WP      ADO50281_07        700001    810000
WP      ADO50281_08        800001    910000
WP      ADO50281_09        900001   1010000
WP      ADO50281_10       1000001   1110000
WP      ADO50281_11       1100001   1210000
WP      ADO50281_12       1200001   1310000
WP      ADO50281_13       1300001   1410000
WP      ADO50281_14       1400001   1510000
WP      ADO50281_15       1500001   1610000
WP      ADO50281_16       1600001   1691134
ID   ADO50281 standard; DNA; 1691134 BP.
XX
AC   ADO50281;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human phosphodiesterase 4D (PDE4D) gene.
XX
KW   ds; stroke; phosphodiesterase 4D; PDE4D; human; gene; SNP;
KW   single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(135641,a)
FT                   /*tag=  a
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(732790,t)
FT                   /*tag=  b
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(735966,c)
FT                   /*tag=  c
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(736516,c)
FT                   /*tag=  d
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(850001,c)
FT                   /*tag=  e
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(852776,a)
FT                   /*tag=  f
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(853575,c)
FT                   /*tag=  g
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(856468,c)
FT                   /*tag=  h
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(870924,c)
FT                   /*tag=  i
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1027643,t)
FT                   /*tag=  j
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1044016,c)
FT                   /*tag=  k
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1254849,c)
FT                   /*tag=  l
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1258161,c)
FT                   /*tag=  m
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1268007,g)
FT                   /*tag=  n
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1268187,t)
FT                   /*tag=  o
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1268553,c)
FT                   /*tag=  p
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1272669,c)
FT                   /*tag=  q
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1272910,a)
FT                   /*tag=  r
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1273220,c)
FT                   /*tag=  s
```

```
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1273240,c)
FT                   /*tag=  t
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1345286,c)
FT                   /*tag=  u
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1346112,c)
FT                   /*tag=  v
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1354377,c)
FT                   /*tag=  w
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1354554,a)
FT                   /*tag=  x
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1354675,t)
FT                   /*tag=  y
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1357081,a)
FT                   /*tag=  z
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1363021,c)
FT                   /*tag=  aa
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1373349,c)
FT                   /*tag=  ab
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1373384,g)
FT                   /*tag=  ac
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1376149,g)
FT                   /*tag=  ad
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1385445,c)
FT                   /*tag=  ae
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1409210,a)
FT                   /*tag=  af
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1414804,t)
FT                   /*tag=  ag
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1428284,c)
FT                   /*tag=  ah
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1431800,t)
FT                   /*tag=  ai
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1449904,a)
FT                   /*tag=  aj
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1574301,c)
FT                   /*tag=  ak
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1574615,c)
FT                   /*tag=  al
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1575634,a)
FT                   /*tag=  am
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1580088,g)
FT                   /*tag=  an
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1581078,a)
FT                   /*tag=  ao
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1582418,t)
FT                   /*tag=  ap
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1590837,a)
FT                   /*tag=  aq
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1590936,a)
FT                   /*tag=  ar
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1591011,g)
FT                   /*tag=  as
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1591047,t)
FT                   /*tag=  at
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1591306,a)
FT                   /*tag=  au
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1591583,c)
FT                   /*tag=  av
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1594994,g)
FT                   /*tag=  aw
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1636902,c)
FT                   /*tag=  ax
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1638550,c)
FT                   /*tag=  ay
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1640663,c)
FT                   /*tag=  az
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1641954,t)
FT                   /*tag=  ba
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(1655748,a)
FT                   /*tag=  bb
FT                   /note= "Single nucleotide polymorphism"
XX
PN   US2004091865-A1.
XX
PD   13-MAY-2004.
XX

  Query Match             21.8%;  Score 435.6;  DB 12;  Length 110000;
  Best Local Similarity   67.1%;  Pred. No. 3.8e-94;
  Matches  803;  Conservative    0;  Mismatches  290;  Indels  104;  Gaps    9;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              || ||||||||  | ||||||||||||||||||||   |||||||||| ||||||||||
Db      99343 TCTGGTGGGGAGGTAGAGAACTTTTGTGTCTAGCTCGGGGATTGTAAACGCACCAATCAG 99284

Qy        491 CACTCTGT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAA 542
              ||| ||||         | | |       |   |||   ||||||| || |||| ||||
Db      99283 CACCCTGTCAAAACAGACCAATCAGCTCTCTGTAAAACAGACCAATCGGCTCTCTTTAAA 99224

Qy        543 ATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGA 602
              |||||||||||||||||| ||||| |||| || ||||| || ||||| | |||  ||| |
Db      99223 ATGGACCAATCAGCAGGACGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCAA 99164

Qy        603 GCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCT 662
              |||||||||| |||| ||  ||| ||||| ||  |||||| |||||||||||||  ||
Db      99163 GCCAGCAGTGGGAACCTGCTGAGGTTCCCTTTCATGCTGTGGTAGCTTTGTTCTTTCACT 99104

Qy        663 CTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTA 722
              |||  |||||||||||| || |||||| |||||| ||||||||| ||||||||||||||||
Db      99103 CTTTGCAATAAATCTTGTTGTTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTA 99044

Qy        723 ACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTG 781
              ||||||||   || ||||||  ||||||  |||||||  | | | ||| ||||||| | |
Db      99043 ACACTCACNATGAAGGTCTGCAGCTTCACCCCTGAAGCTAGCAAAACCGCGAACCCGCCG 98984

Qy        782 GAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCT 841
              | |||||| || |||||||||  || || | ||||||||||||||||||||||||||| |
Db      98983 GGAGGAACGAACAACTCCCGACATG-TGGCCTTAAGAGCTGTAACACTCACTGCGAAGGT 98925

Qy        842 CTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGG 901
              |||||||||| ||||| || ||| ||||||   |||||||||||| |||||||||
```

```
Db      98924 TTGCAGCTTCATTCCTG-AGCCAGCGAGACCGTGAACCCACCAGAAGAAAGAAACTCCGA 98866

Qy        902 ACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACAC 961

Db      98865 ACACATCCAAACATCAGAAGGAACAAACTCCAGACACGCTGGCTTTAAGAACTGTAACAT 98806

Qy        962 TCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAG 1021

Db      98805 TCACTGAGAGGGTCTGCGGCTTCATTCTTGAAATCAGTAAGACCAAGAACCCACC----- 98751

Qy       1022 GAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACT 1081

Db      98750 -----AACTCAGGACACAGTGTGA---------------------------------- 98732

Qy       1082 CCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACA 1141

Db      98731 -------------------------TTTTCACAGCCGGAAATTGACTTGGAAATCCTTA 98698

Qy       1142 TGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAG 1201

Db      98697 AGTTACTTTTTTTTTAATGCAGCTTTTGGATAGCCATAAAAATCTTATACCTTCCTTTAA 98638

Qy       1202 CTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTC 1261

Db      98637 TGTTATTTAAATTTTTGAAATAAGTTCAGTAACAAGAAAACAAGATGGACACAA------ 98584

Qy       1262 TTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCG 1321

Db      98583 -CAGTAACTTTAAAACGATTTTCTTTAGATCTTAAAAGATACGTATTTCAGCCAGGCATG 98525

Qy       1322 ATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCA 1381

Db      98524 GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCACGAGGTCA 98465

Qy       1382 GGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACA---- 1437

Db      98464 GGAGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGCCTCTAGTAAAAATACAAAAA 98405

Qy       1438 ---------AAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAG 1487

Db      98404 AAAAAAAAAAAAATTAGCCCGGAGTGGGGGCAGGCGCCTGTAGTCCTGGCTACTCGGGAG 98345

Qy       1488 GCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACA 1547

Db      98344 GCTGAGGCAGGAGAATAGCATGAACCCAGGAGGCGGAGCTTGCAATAAGCCGAGATTGTG 98285

Qy       1548 CCACAGCACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAA 1604

Db      98284 CCATTGCACTCCAGCCTGGGTGACAGTGTGAGACTCCATCTCAAAAAAAAAAAGATA 98228


RESULT 26
AEB85185_00/c
WP Sequence split into 17 fragments  LOCUS AEB85185 Accession Aeb85185
WP      Fragment Name       Begin      End
WP      AEB85185_00             1    110000
WP      AEB85185_01        100001    210000
WP      AEB85185_02        200001    310000
WP      AEB85185_03        300001    410000
WP      AEB85185_04        400001    510000
WP      AEB85185_05        500001    610000
WP      AEB85185_06        600001    710000
WP      AEB85185_07        700001    810000
WP      AEB85185_08        800001    910000
WP      AEB85185_09        900001   1010000
WP      AEB85185_10       1000001   1110000
WP      AEB85185_11       1100001   1210000
WP      AEB85185_12       1200001   1310000
WP      AEB85185_13       1300001   1410000
WP      AEB85185_14       1400001   1510000
WP      AEB85185_15       1500001   1610000
WP      AEB85185_16       1600001   1691140
ID   AEB85185 standard; DNA; 1691140 BP.
XX
AC   AEB85185;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Human phosphodiesterase 4D gene SEQ ID NO:1.
XX
KW   diagnosis; cerebrovascular ischemia; phosphodiesterase 4D; chromosome-5;
KW   cerebroprotective; vasotropic; gene; ds; single nucleotide polymorphism;
KW   SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       78552
FT                   /*tag=  a
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       80127
FT                   /*tag=  b
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       82244
FT                   /*tag=  c
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       83119
FT                   /*tag=  d
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       88456
FT                   /*tag=  e
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       88614
FT                   /*tag=  f
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       90541
FT                   /*tag=  g
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       91022
FT                   /*tag=  h
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       replace(91470,A)
FT                   /*tag=  i
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT   variation       92148
FT                   /*tag=  j
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       99035
FT                   /*tag=  k
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       100795
FT                   /*tag=  l
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       102977
FT                   /*tag=  m
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       104552
FT                   /*tag=  n
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       105225
FT                   /*tag=  o
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
```

```
FT   variation       105792
FT                   /*tag=  p
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       107317
FT                   /*tag=  q
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       107849
FT                   /*tag=  r
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   exon            108127. .108217
FT                   /*tag=  s
FT                   /number= 1 - 4D7-A
FT                   /note= "see page 47 Table 10A"
FT   intron          108218. .142206
FT                   /*tag=  t
FT   variation       109301
FT                   /*tag=  u
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       111252
FT                   /*tag=  v
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       111781
FT                   /*tag=  w
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       118914
FT                   /*tag=  x
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       replace(120628,R)
FT                   /*tag=  y
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "given as N in the sequence"
FT   variation       123312
FT                   /*tag=  z
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       123426
FT                   /*tag=  aa
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       125304
FT                   /*tag=  ab
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       129360
FT                   /*tag=  ac
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       replace(129361,Y)
FT                   /*tag=  ad
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "given as N in the sequence"
FT   variation       130391
FT                   /*tag=  ae
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       replace(131865,M)
FT                   /*tag=  af
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "given as N in the sequence"
FT   variation       132562
FT                   /*tag=  ag
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       133371
FT                   /*tag=  ah
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       135112
FT                   /*tag=  ai
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       135152
FT                   /*tag=  aj
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       replace(135641,A)
FT                   /*tag=  ak
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT   variation       136265
FT                   /*tag=  al
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       138456
FT                   /*tag=  am
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   exon            142207. .142328
FT                   /*tag=  an
FT                   /number= 2 - PDE4D7-1
FT                   /note= "see page 47 Table 10A"
FT   intron          142329. .257649
FT                   /*tag=  ao
FT   variation       142556
FT                   /*tag=  ap
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"
FT   variation       142757
FT                   /*tag=  aq
FT                   /standard_name= "single nucleotide polymorphism (SNP)"
FT                   /note= "no variation given - see Table 2C"

  Query Match             21.8%;  Score 435.6;  DB 14;  Length 110000;
  Best Local Similarity   67.1%;  Pred. No. 3.8e-94;
  Matches  803;  Conservative    0;  Mismatches  290;  Indels  104;  Gaps    9;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              || ||||||||  | ||||||||||||||||||||    |||||||||| ||||||||||
Db      99343 TCTGGTGGGGAGGTAGAGAACTTTTGTGTCTAGCTCGGGGATTGTAAACGCACCAATCAG 99284

Qy        491 CACTCTGT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAA 542
              |||  ||||        | | |      |   |||    ||||||| || |||| ||||
Db      99283 CACCCTGTCAAAACAGACCAATCAGCTCTCTGTAAAACAGACCAATCGGCTCTCTTTAAA 99224

Qy        543 ATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGA 602
              |||||||||||||||||| ||||| |||| || ||||| || ||||| | |||  ||| |
Db      99223 ATGGACCAATCAGCAGGACGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCAA 99164

Qy        603 GCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCT 662
              |||||||||| ||||  ||  ||| ||||| ||  |||||| |||||||||||  ||
Db      99163 GCCAGCAGTGGGAACCTGCTGAGGTTCCCTTTCATGCTGTGGTAGCTTTGTTCTTTCACT 99104

Qy        663 CTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTA 722
              |||  |||||||||||| || |||||| |||||| ||||||||| ||||||||||||||
Db      99103 CTTTGCAATAAATCTTGTTGTTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTA 99044

Qy        723 ACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTG 781
              ||||||||   || ||||||  ||||||  |||||||  | | | ||| |||||||| | |
Db      99043 ACACTCACNATGAAGGTCTGCAGCTTCACCCCTGAAGCTAGCAAAACCGCGAACCCGCCG 98984

Qy        782 GAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCT 841
              | |||||| || |||||||||  || || | |||||||||||||||||||||||||| |
Db      98983 GGAGGAACGAACAACTCCCGACATG-TGGCCTTAAGAGCTGTAACACTCACTGCGAAGGT 98925

Qy        842 CTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGG 901
              |||||||||| ||||| || ||| ||||||   |||||||||||||| |||||||||  |
Db      98924 TTGCAGCTTCATTCCTG-AGCCAGCGAGACCGTGAACCCACCAGAAGAAAGAAACTCCGA 98866

Qy        902 ACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACAC 961
              ||||| |  || ||| |||||||||||||||||||| |   |||| ||| ||||||||
```

```
Db      98865 ACACATCCAAACATCAGAAGGAACAAACTCCAGACACGCTGGCTTTAAGAACTGTAACAT 98806

Qy        962 TCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAG 1021

Db      98805 TCACTGAGAGGGTCTGCGGCTTCATTCTTGAAATCAGTAAGACCAAGAACCCACC----- 98751

Qy       1022 GAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACT 1081

Db      98750 -----AACTCAGGACACAGTGTGA---------------------------------- 98732

Qy       1082 CCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACA 1141

Db      98731 --------------------------TTTTCACAGCCGGAAATTGACTTGGAAATCCTTA 98698

Qy       1142 TGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAG 1201

Db      98697 AGTTACTTTTTTTTTAATGCAGCTTTTGGATAGCCATAAAAATCTTATACCTTCCTTTAA 98638

Qy       1202 CTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTC 1261

Db      98637 TGTTATTTAAATTTTTGAAATAAGTTCAGTAACAAGAAAACAAGATGGACACAA------ 98584

Qy       1262 TTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCG 1321

Db      98583 -CAGTAACTTTAAAACGATTTTCTTTAGATCTTAAAAGATACGTATTTCAGCCAGGCATG 98525

Qy       1322 ATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCA 1381

Db      98524 GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCACGAGGTCA 98465

Qy       1382 GGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACA---- 1437

Db      98464 GGAGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGCCTCTAGTAAAAATACAAAAA 98405

Qy       1438 ---------AAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAG 1487

Db      98404 AAAAAAAAAAAAATTAGCCCGGAGTGGGGGCAGGCGCCTGTAGTCCTGGCTACTCGGGAG 98345

Qy       1488 GCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACA 1547

Db      98344 GCTGAGGCAGGAGAATAGCATGAACCCAGGAGGCGGAGCTTGCAATAAGCCGAGATTGTG 98285

Qy       1548 CCACAGCACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAA 1604

Db      98284 CCATTGCACTCCAGCCTGGGTGACAGTGTGAGACTCCATCTCAAAAAAAAAAAGATA 98228
```

RESULT 27
ADL27128
ID ADL27128 standard; DNA; 28670 BP.
XX
AC ADL27128;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human genomic sequence for IQGAP1.
XX
KW Human; ds; gene; CA gene; carcinoma associated gene; cytostatic; cancer;
KW carcinoma; lymphoma.
XX
OS Homo sapiens.
XX
PN US2003216558-A1.
XX
PD 20-NOV-2003.
XX
PF 20-DEC-2001; 2001US-00034650.
XX
PR 22-DEC-2000; 2000US-00747377.
PR 02-MAR-2001; 2001US-00798586.
XX
PA (MORR/) MORRIS D W.
PA (ENGE/) ENGELHARD E K.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2003-902052/82.
XX
PT New carcinoma associated gene, useful for preparing a composition for
PT diagnosing or treating carcinoma.
XX
PS Claim 1; SEQ ID NO 16; 342pp; English.
XX
CC The invention relates to a new recombinant nucleic acid (from a CA,
CC carcinoma associated, gene) appearing as ADL27113-ADL27172. Also included
CC are a host cell comprising the recombinant nucleic acid or expression
CC vector, an expression vector comprising the recombinant nucleic acid, a
CC recombinant protein (a carcinoma associated protein) comprising the
CC sequence encoded by the nucleic acid, a method for screening drug
CC candidates, a method for screening for a bioactive agent capable of
CC binding to (or modulating the activity of) a carcinoma associated
CC protein, a method for evaluating the effect of a candidate carcinoma
CC drug, a method of diagnosing carcinoma, a method for inhibiting the
CC activity of a carcinoma associated protein, a method of treating
CC carcinomas, a method of neutralising the effect of a carcinoma associated
CC protein, a biochip comprising one or more nucleic acid segments of the
CC nucleic acid, a method of diagnosing carcinoma or propensity to carcinoma
CC and a method of determining carcinoma associated gene copy number. The
CC nucleic acid is useful for preparing a composition for diagnosing or
CC treating carcinoma especially lymphomas. The present sequence is the
CC genomic sequence from a human carcinoma associated gene.
XX
SQ Sequence 28670 BP; 7471 A; 6745 C; 6201 G; 8252 T; 0 U; 1 Other;

Query Match 21.7%; Score 433.4; DB 11; Length 28670;
Best Local Similarity 88.6%; Pred. No. 8.3e-94;
Matches 504; Conservative 0; Mismatches 61; Indels 4; Gaps 3;

```
Qy        476 AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535

Db      23250 AAACGCACCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAACGCACCAGTCAGCACT 23309

Qy        536 CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGC 595

Db      23310 CTGTAAAATGGACTAATCAGCAGGATGTGGGTGGGGCCAAATAAGGCAATAAAAGCTGGC 23369

Qy        596 CACCCGAGCCAGCAG-TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTT 654

Db      23370 CAGCGAAGCCAGCAGCAGGCAACCCACTCGGGTGCCCTTCCATGCTGTGGAAGCTTTGTT 23429

Qy        655 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTAT 714

Db      23430 CTTTCGCTCTTCACAGTAAATCTTGCTGCTGCTCA--CTCTGGGTCCGCACTACCTTCAT 23487

Qy        715 GAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACA-GACCACGA 773

Db      23488 GAGCTGTAACACTCAATGCGAGGGTCTGCGGCTGCATTCCTGAAGTCACCAAGACCACAA 23547

Qy        774 ACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACT 833

Db      23548 ACCCACCGGGAGGAACAAACAACTCCAGATGAGCCACCTTTAAGAGCTGTAACACTCACT 23607

Qy        834 GCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGA 893

Db      23608 GCGAAGGTGTGCGGCTTCACTCCTGAAGTCAGTGACACCACGAACCCACTGGAAGGAAGA 23667

Qy        894 AACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGC 953

Db      23668 AACTCCGGACACATCTGAACATCTGAAGGAACAAACTCCGGACACACCATCTTTAAGAGC 23727

Qy        954 TGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCC 1013

Db      23728 TGTAACGCTCACCGTGAAGGTCCGTGGTTTCATTCTTGAAGTCAGCGAGACCAAGCATCC 23787

Qy       1014 ACCGGAAGGAACAAATTCCAGACACAGTA 1042
```

```
Db      23788 ACCAAAAGGAATAAATTCCAGACCCACTA 23816


RESULT 28
ADQ20461/c
ID   ADQ20461 standard; DNA; 166181 BP.
XX
AC   ADQ20461;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 3281.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 3281; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 166181 BP; 37688 A; 45326 C; 45775 G; 37392 T; 0 U; 0 Other;

  Query Match             21.2%;  Score 425.2;  DB 12;  Length 166181;
  Best Local Similarity   66.2%;  Pred. No. 1.4e-91;
  Matches  797;  Conservative    0;  Mismatches  338;  Indels   69;  Gaps   10;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||||||  ||||||||  ||||||||  ||||  |  ||||||||||||||||||||||
Db      91497 TCTAGTGGGGATGTGGAGAACCTTTGTGTCAAGCTCAGGGATTGTAAATGCACCAATCAG 91438

Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCA 550
              | | ||||                  |||   |||| || ||  ||||||||||||||||
Db      91437 CGCCCTGTCA----------------AAACAGACCACTCGGCTCTCTGTAAAATGGACCA 91394

Qy        551 ATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAG 610
              ||||||||||||||| |||| || | ||| || ||||| | |||  ||| |||||||
Db      91393 ATCAGCAGGATGTGGGTGGGGCCAGACAAGAGAATAAAAGCAGGCTGCCCCAGCCAGCTC 91334

Qy        611 TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAA 670
              |||||||||  |||||||||  ||||||  ||||||||||||||||||||  ||||||  |||

Db      91333 TGGCAACCCGCTCGGGTCCTCTTCCATGCTGTGGAAGCTTTGTTCTTTCGCTCTTTGCAA 91274

Qy        671 TAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCAC 730
              ||||||||||||||||||| |||||| |||||||||  ||||||||||||||||||  |||||
Db      91273 TAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTAACGCTCAC 91214

Qy        731 TGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCACGAACCCACTGGAAGGAAC 789
              |||  ||||||  ||||||  ||||||||  ||  |||||||  |||||||| ||  |||||
Db      91213 CACGAAGGTCTGCAGCTTCACTCCTGAAGCCATCCAGACCATGAACCCACAGGGAGGAAA 91154

Qy        790 AAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCT 849
              ||  ||||  |  ||  |  |   ||  ||||||||  ||||||||||   |  |||  |||||||||
Db      91153 GAACAACTGCGGACGCGTCTCC-TTAAGAGCCGTAACACTCACCACAAAGGTCTGCAGCT 91095

Qy        850 TCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCT 909
              |||||||||  ||  |||  ||||||  |  ||||||||||||||||||||||||   |||||  |
Db      91094 TCACTCCTG-AGCCAGCGAGACCGCGAACCCACCAGAAGGAAGAAACTCCAAACACATCC 91036

Qy        910 GAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCA 969
              |||  |||  ||||||||||||||||||||  ||  ||||  |||  ||||||     |  |||
Db      91035 GAACATCAGAAGGAACAAACTCCAGACACGCCGCCTTTAAGAACTGTAA----TATCGCG 90980

Qy        970 AGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC---GGAAGGAACA 1026
              ||||||||  ||||||||||||||||||  ||||||||||||||||     |  |||
Db      90979 AGGGTCTGCCGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCAATTCCGGACACA 90920

Qy       1027 AATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACTCCAGT 1086
              |      ||  ||  |    |  ||||  |  |       |  |  ||  ||    |    |
Db      90919 ATACTTTGAAACCATGACTATCCTGTTTGCTGACAAATTATCACCCTAGGGATTTTGCAT 90860

Qy       1087 CATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACATGTGG 1146
              |    |    | |||  | |    |    |    | | ||  |    |    | | |
Db      90859 CCGCCAGTGATTCATGCCTGAACCAGTTATTGCTACAGTGCCTGCAACAGTCTCTTAACT 90800

Qy       1147 GAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCTGGA 1206
              |  || | ||   ||  |      ||| |  |       | ||   | | | |  |
Db      90799 GGTTTCTCATCCTTCAACCCCTCCCTGCTCCAATTCATCCTTTG---CATCTTTGATCTA 90743

Qy       1207 AGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTTGC 1266
              | |    |  ||  | |  | |  | |   |||  |   ||  |   ||  ||
Db      90742 GGGAACTTCCTGAGGCACAGGTTTAATTACATGACTTTCCTCCTCAAAAACCTTCGGTGG 90683

Qy       1267 ATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCGATGGC 1326
              ||       ||  ||||  |     | |  |          ||  ||||||     ||||
Db      90682 CTCCCTACTATTTAGAGAAGTCCAAACTCCTT------GGCTAAGGCCAGGCATAGTGGC 90629

Qy       1327 TCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCAGGCGT 1386
              |||  ||||||||||                                  | ||||  |
Db      90628 TCACTCCTGTAATCCC--------------------------------CAGCCAGGAAT 90602

Qy       1387 TCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGC 1446
              | ||||| |||| |||||||||||||||||| |||||||||||||||||| ||||||||
Db      90601 TTGAGACAAGCCCGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACACAAATTAGC 90542

Qy       1447 CTGGTGTGGTGGCACGCATCTGTAATCCCA-GTACTCAGGAGGCTGAGGCAGGGGAATCT 1505
              |  ||||||||||  |   ||||||||||| ||||| ||||||||||||||| ||||||
Db      90541 CACGTGTGGTGGCGGGTGCCTGTAATCCCAGGTACTTAGGAGGCTGAGGCAGGAGAATCT 90482

Qy       1506 CTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCATCCTG 1565
              |||||||  ||||||  |||||||||||||||||||||  ||||||  ||||||||  |||
Db      90481 CTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCAAGAGGGCACCACTGCACTCCAGTCTG 90422

Qy       1566 GGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAATATAAT 1625
              ||  |||| | ||||||||      |||| || || |  || |  ||    ||  | |
Db      90421 GGCAACAGTGTGAGACTCTTATCTCAAAATAATAATAGTAATAATAATAATAATAACAGG 90362

Qy       1626 CAAG 1629
              | ||
Db      90361 CTAG 90358


RESULT 29
ADQ18633/c
```

```
ID   ADQ18633 standard; DNA; 166181 BP.
XX
AC   ADQ18633;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1452.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1452; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 166181 BP; 37688 A; 45326 C; 45775 G; 37392 T; 0 U; 0 Other;

  Query Match             21.2%;  Score 425.2;  DB 12;  Length 166181;
  Best Local Similarity   66.2%;  Pred. No. 1.4e-91;
  Matches  797;  Conservative    0;  Mismatches  338;  Indels   69;  Gaps   10;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||||||  ||||||||  |||||||| |||| | ||||||||||||||||||||
Db      91497 TCTAGTGGGGATGTGGAGAACCTTTGTGTCAAGCTCAGGGATTGTAAATGCACCAATCAG 91438

Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCA 550
              | | ||||                  |||   |||| || || ||||||||||||||||||
Db      91437 CGCCCTGTCA----------------AAACAGACCACTCGGCTCTCTGTAAAATGGACCA 91394

Qy        551 ATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAG 610
              ||||||||||||||||| |||| || | ||| || ||||| | |||  ||| |||||||
Db      91393 ATCAGCAGGATGTGGGTGGGGCCAGACAAGAGAATAAAAGCAGGCTGCCCCAGCCAGCTC 91334

Qy        611 TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAA 670
              |||||||| |||||||| ||||||  |||||||||||||||||||| ||||||  |||
Db      91333 TGGCAACCCGCTCGGGTCCTCTTCCATGCTGTGGAAGCTTTGTTCTTTCGCTCTTTGCAA 91274

Qy        671 TAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCAC 730
              |||||||||||||||||| |||||| |||||||| |||||||||||||||||| |||||
Db      91273 TAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTAACGCTCAC 91214

Qy        731 TGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCACGAACCCACTGGAAGGAAC 789
              ||| ||||||  |||||| |||||||| ||  ||||||| |||||||| || |||||
Db      91213 CACGAAGGTCTGCAGCTTCACTCCTGAAGCCATCCAGACCATGAACCCACAGGGAGGAAA 91154

Qy        790 AAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCT 849
              || |||| | || | |   || |||||||| ||||||||||  | ||| ||||||||||
Db      91153 GAACAACTGCGGACGCGTCTCC-TTAAGAGCCGTAACACTCACCACAAAGGTCTGCAGCT 91095

Qy        850 TCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCT 909
              ||||||||| || ||| |||||| | ||||||||||||||||||||||||  |||||
Db      91094 TCACTCCTG-AGCCAGCGAGACCGCGAACCCACCAGAAGGAAGAAACTCCAAACACATCC 91036

Qy        910 GAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCA 969
              ||| ||| |||||||||||||||||||| ||  |||| ||| ||||||     | |||
Db      91035 GAACATCAGAAGGAACAAACTCCAGACACGCCGCCTTTAAGAACTGTAA----TATCGCG 90980

Qy        970 AGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC---GGAAGGAACA 1026
              |||||||| |||||||||||||||| ||||||||||||||||        | |||
Db      90979 AGGGTCTGCCGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCAATTCCGGACACA 90920

Qy       1027 AATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACTCCAGT 1086
              |     ||  ||   |   | |||| |  |     |  |  ||  ||    |    |
Db      90919 ATACTTTGAAACCATGACTATCCTGTTTGCTGACAAATTATCACCCTAGGGATTTTGCAT 90860

Qy       1087 CATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACATGTGG 1146
              |    |   | |||  | |   |  |    | | ||  |  |   | | |
Db      90859 CCGCCAGTGATTCATGCCTGAACCAGTTATTGCTACAGTGCCTGCAACAGTCTCTTAACT 90800

Qy       1147 GAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCTGGA 1206
              |  || | ||    ||  |     ||| |  |       | ||    | | | | |  |
Db      90799 GGTTTCTCATCCTTCAACCCCTCCCTGCTCCAATTCATCCTTTG---CATCTTTGATCTA 90743

Qy       1207 AGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTTGC 1266
              | |   |  ||  | |  | |   |||  |  ||  |  ||  ||
Db      90742 GGGAACTTCCTGAGGCACAGGTTTAATTACATGACTTTCCTCCTCAAAAACCTTCGGTGG 90683

Qy       1267 ATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCGATGGC 1326
              ||      || |||| |      | |  |        ||  ||||||     ||||
Db      90682 CTCCCTACTATTTAGAGAAGTCCAAACTCCTT------GGCTAAGGCCAGGCATAGTGGC 90629

Qy       1327 TCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCAGGCGT 1386
              |||  ||||||||||                                    | |||  |
Db      90628 TCACTCCTGTAATCCC--------------------------------CAGCCAGGAAT 90602

Qy       1387 TCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGC 1446
              | ||||| |||| |||||||||||||||||| |||||||||||||||||| ||||||||
Db      90601 TTGAGACAAGCCCGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACACAAATTAGC 90542

Qy       1447 CTGGTGTGGTGGCACGCATCTGTAATCCCA-GTACTCAGGAGGCTGAGGCAGGGGAATCT 1505
              |  |||||||||| |  |||||||||| ||||| |||||||||||||||| ||||||
Db      90541 CACGTGTGGTGGCGGGTGCCTGTAATCCCAGGTACTTAGGAGGCTGAGGCAGGAGAATCT 90482

Qy       1506 CTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCATCCTG 1565
              |||||||  |||||| |||||||||||||||||||||  |||||| ||||||| |||
Db      90481 CTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCAAGAGGGCACCACTGCACTCCAGTCTG 90422

Qy       1566 GGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAATATAAT 1625
              ||  |||| | |||||||||      |||| || || |  || |  ||   ||  | |
Db      90421 GGCAACAGTGTGAGACTCTTATCTCAAAATAATAATAGTAATAATAATAATAATAACAGG 90362

Qy       1626 CAAG 1629
              | ||
Db      90361 CTAG 90358


RESULT 30
ACN44046/c
ID   ACN44046 standard; DNA; 260027 BP.
XX
AC   ACN44046;
XX
DT   18-NOV-2004  (first entry)
```

```
XX
DE   Human genomic sequence hCG1735292.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 298; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 260027 BP; 72144 A; 50369 C; 51877 G; 83517 T; 0 U; 2120 Other;

  Query Match             21.2%;  Score 424.6;  DB 11;  Length 260027;
  Best Local Similarity   66.0%;  Pred. No. 2.3e-91;
  Matches  714;  Conservative    0;  Mismatches  344;  Indels   23;  Gaps    6;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              || | |||   || ||   |||||||||| |||||||||||||||||||||||||||||
Db     177582 AACGCCTTTCAAACTCTTTCAATGCCTTTCCAACTCTTATACCAACCTCTGGAGTTGGGG 177523

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||| ||| |||||||||||||||||||||||||||||||| || |||| ||||||
Db     177522 AACATGGCATCTTCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTATTACTCGCCTTTGG 177463

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||| ||| | |||||||||||||||||| ||||||||||||||||||
Db     177462 GCCCTGTATTTTTAATCTCCTTGTCAAATTTGTTTCCTCTACGATCGAGGCCATCAAGCT 177403

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||| |||||||||||||||||| |  ||||||
Db     177402 ACAGATGGTCTTACAAATGGAACCCCAAATAAGCTCAACTAACAACTTCTACCAAGGACC 177343

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              |||||| |||||| |||||||||| | ||||||| | || ||||||||||||||||||
Db     177342 CCTGGATCGACCCACTGGCCCTTTGACTGGCCTAGACAGTTCCCCTCTGGAGGACACTAC 177283

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              |||||||||| ||||||||  ||| ||||||| |||||||||| |  | |||| | ||
Db     177282 CACTGCAGGGTCCCTTCTTTGCCCTTATCCAGTAGGAAGTAGCTAGAATGATCATTGCCC 177223

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||||||||| |||||||||| ||| |||||||||||||||| ||||||||||

Db     177222 AATTCCCAACAGCAGTTGGGGTGTCCTATTTAGAGGGGGGATTGAGAGGGGAAGCCAGCT 177163

Qy        420 GGGCTT-CTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
              ||  || ||||||   |||| |||||||||||||||| |||||||||| |||| ||||||
Db     177162 GGATTTCCTGGGTTGAGTGGAGACTTGGAGAACTTTTCTGTCTAGCTAGAGGACTGTAAA 177103

Qy        479 TGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTG 538
              ||||||||||  ||||||| |||||||||||||||||||||| ||||||||||||||
Db     177102 CACACCAATCAGTGCTCTGTGCCTAGCTAAAGGATTGTAAATGCAACAATCAGCACTCTG 177043

Qy        539 TAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCAC 598
              |||||  |  |          |  | | ||  |  | |     | |   | |
Db     177042 TAAAAATGCACCCATCAGCGCTCTGTATCTAGCTAAAGGATTGTAAATGCACCAATCAGC 176983

Qy        599 CCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTT 658
              |  |  |  |  |   ||| |   | |  |  |  |   ||   | |   |  |
Db     176982 ACTCGGTAAAAACG--CACCAATCAGCGCTCTGTGTCTAGCTAAAGGATTGTAAATGCAC 176925

Qy        659 TGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGC 718
              |  ||  |||  |  |          |  |||  |  |||||  |     |    |  |
Db     176924 TAATCAGCACTCTGTAAAAATGCACCAGTCAG-CACTGTGTCTAGCTAAAGGATTGTAAA 176866

Qy        719 TGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCA 778
              || |||| |||  |  |   |||   ||   |  |  | |  |||||  |
Db     176865 TGCAACAATCAGCACTCTGTAAAATGGACCAATCAGCGCTCTGTAAAATGGACGAATCAG 176806

Qy        779 CTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAA 838
              | |||      |   ||   || |    | |  |||       | |     | ||  |
Db     176805 CAGGACATGGCGGGGA---GCCAAATAACGGAATAAAAGCTAGCCACCCAAGCCAGCAGA 176749

Qy        839 GCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTC 898
              |      ||  || |     | |    ||  |      |  |       |  || ||
Db     176748 GGCAACTAGGGTCCCCTTCCACGCTGTGGAAGCTTTGTTCTTTCGCTCTTAATAAATCTT 176689

Qy        899 TGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAA 958
              |   |              ||    |     | ||||| |||| ||||||||||
Db     176688 GCTGCTGTTC--------------ACTCTTTGGGTCCGCACCACCTTTAAGAGCTGTAA 176644

Qy        959 CACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGG 1018
              |||||||  | | ||||| |||||||||||||| |||||||  ||||||||||||||| |
Db     176643 CACTCACAACGAAGGTCTATGGCTTCATTCTTCAAGTCAGAGAGACCAAGAACCCACCAG 176584

Qy       1019 AAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTT 1078
              ||||||| || ||  ||||||  | |||  ||  |  ||  ||  |   |  | || |
Db     176583 AAGGAACCAACTCTGGACACACAATGAAGGCTTGACCATTTTTCAATATGCTTAAAGGCT 176524

Qy       1079 A 1079
              |
Db     176523 A 176523


RESULT 31
ADL13904
ID   ADL13904 standard; DNA; 164772 BP.
XX
AC   ADL13904;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Osteoarthritis-associated polymorphic nucleotide #436.
XX
KW   ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW   joint space narrowing; osteophyte development; joint pain;
KW   osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO2003054166-A2.
XX
PD   03-JUL-2003.
XX
PF   19-DEC-2002; 2002WO-US041225.
XX
```

```
PR   20-DEC-2001; 2001US-0342603P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA,  Schafer A;
XX
DR   WPI; 2003-559141/52.
XX
PT   Determining susceptibility of an individual to joint space narrowing,
PT   osteophyte development and/or joint pain comprises identifying whether
PT   the individual has at least one polymorphism in a polynucleotide encoding
PT   a protein.
XX
PS   Disclosure; SEQ ID NO 436; 297pp; English.
XX
CC   The invention relates to a method of determining susceptibility of an
CC   individual to joint space narrowing and/or osteophyte development and/or
CC   joint pain comprising identifying whether the individual has at least one
CC   polymorphism in a polynucleotide encoding at least one of the protein
CC   listed in the specification. The methods, composition and agent are
CC   useful for modulating the susceptibility of an individual to joint space
CC   narrowing and/or osteophyte development and/or joint pain that is
CC   associated with a disease, preferably osteoarthritis. The cell line and
CC   the non-human animal are useful for screening for an agent for diagnosing
CC   an individual having susceptibility to joint space narrowing and/or
CC   osteophyte development and/or joint pain. This sequence corresponds to
CC   the polynucleotide encoding a protein listed in the specification. (Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences).
XX
SQ   Sequence 164772 BP; 50645 A; 32137 C; 31960 G; 50022 T; 0 U; 8 Other;

  Query Match             21.0%;  Score 420.4;  DB 10;  Length 164772;
  Best Local Similarity   78.7%;  Pred. No. 2e-90;
  Matches  573;  Conservative    0;  Mismatches   81;  Indels   74;  Gaps    3;

Qy        387 GTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGG 446
              |||  | |      | ||||||||||| |   |||||||||||||||||||||||||||||
Db      12172 GTTCAGGGTTCCCAGTGAGAGGTGAATCTGTCTGGGCTTCTGGGTCAGGTGGGGACTTGG 12231

Qy        447 AGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCT 506
              ||||||||| || |||||||||| |||||||||||| ||||||| ||||||||||||||
Db      12232 AGAACTTTTCTGCCTAGCTAAAGGGTTGTAAATGCACAAATCAGCGCTCTGTGTCTAGCT 12291

Qy        507 AAAGGATTGTAAATGCACCAATCAGCACTCTGT-------------------------- 539
              ||||| ||||||| |||||||||||||||| |||
Db      12292 AAAGGTTTGTAAACGCACCAATCAGCACTTTGTAAAAAACACACCAATCAACGTCTGTGT 12351

Qy        540 -----------------------------------------AAAATGGACCAATCAGCAG 558
                                                        |||| | |||||||||| |
Db      12352 TTAGCTAAAGGTTTTGTAAGTGCACCAATCAGCACTCTGTAAAAACGCACCAATCAGCGG 12411

Qy        559 GATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACC 618
              || |||||| ||| || |||||||| ||||| |||||||| |||||||||| |||||||
Db      12412 GACGTGGGCAGGGCCACATAAGGGAATAAAAGCTGGCCACCTGAGCCAGCAGCGGCAACC 12471

Qy        619 CACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTT 678
                ||  ||||||||||||     |||||||||||||||| |||||| ||||||||||||
Db      12472 TGCTTAGGTCCCCTTCCA-----TGGAAGCTTTGTTCTTTCGCTCTTGACAATAAATCTT 12526

Qy        679 GCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGG 738
              |||||||||||| |||||| |||||||| |||||||||||||||||||||||||||| |
Db      12527 GCTGCTGCTCACTCTTTGGGTCCACACCACCTTTATGAGCTGTAACACTCACTGCGAGTG 12586

Qy        739 TCTGTGGCTTCATTCCTGAAGTCA-ACAGACCACGAACCCACTGGAAGGAACAAAGAACT 797
              |||||||||||||||||||||||   |||||||||||||||  | |||||| || ||||
Db      12587 TCTGTGGCTTCATTCCTGAAGTCAGTGAGACCACGAACCCACCAGGAGGAACGAACAACT 12646

Qy        798 CCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCT 857
              |  || || |  |||||||||||||||||||||||||||  || ||||  ||||||||||
Db      12647 CTGGACGTACCACCTTTAAGAGCTGTAACACTCACTGCAGAGGTCTGTGGCTTCACTCCT 12706

Qy        858 GAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCT 917
              ||||||| |||||||||| |||||| |||||||||||||||||||||| ||||| ||||
Db      12707 TAAGTCAGCGAGACCACAAGCCCACCGGAAGGAAGAAACTCTGGACACATCTGAACATCT 12766

Qy        918 GAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTG 977
              |||||||||||||| |||||||||||||| ||| |||||||||||||| | ||||||||
Db      12767 GAAGGAACAAACTCCGGACACACCATCTTTAAGAACTGTAACACTCACCACGAGGGTCTG 12826

Qy        978 TGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACA 1037
              |||||||||||||||||||  |||||||||||||||| ||||||| ||||||  ||||
Db      12827 CAGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCAGAAGGAATAAATTCTGGACA 12886

Qy       1038 CAGTAGGA 1045
              ||  ||||
Db      12887 CACCAGGA 12894
```

```
RESULT 32
AAZ29173
ID   AAZ29173 standard; DNA; 2834 BP.
XX
AC   AAZ29173;
XX
DT   21-FEB-2000  (first entry)
XX
DE   Targeting sequence-1 homologous to part of 5'non-coding region of G-CSF.
XX
KW   Granulocyte colony stimulating factor; G-CSF; genomic sequence; upstream;
KW   transcription start site; 5'non-coding sequence; DNA construct;
KW   targeting sequence-1; regulatory region; marker gene; selection;
KW   homologous recombination; gene therapy; delivery system; CMV promoter;
KW   haematopoietic progenitor cell; chemotherapy-induced neutropenia;
KW   bone marrow transplantation; congenital neutropenic disorder;
KW   chronic idiopathic disorder; plasmid pGG13; neomycin resistance gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO9957291-A1.
XX
PD   11-NOV-1999.
XX
PF   05-MAY-1999;   99WO-US009924.
XX
PR   07-MAY-1998;   98US-0084649P.
XX
PA   (TRAN-) TRANSKARYOTIC THERAPIES INC.
XX
PI   Treco DA,  Heartlein MW,  Selden RF;
XX
DR   WPI; 2000-072235/06.
XX
PT   Novel genomic sequences used for treating human diseases and disorders.
XX
PS   Disclosure; Fig 6; 58pp; English.
XX
CC   The present DNA sequence is the first targeting sequence that corresponds
CC   to nucleotides 1470-4723 of the human genomic sequence from a region
CC   upstream of the transcription start site of granulocyte colony-
CC   stimulating factor (G-CSF). A DNA construct comprising this targeting
CC   sequence is cloned into the plasmid pGG13, upstream of a CMV promoter and
CC   neomycin resistance gene, that functions as the transcriptional
CC   regulatory sequence and as selectable marker respectively. Homologous
CC   recombination of this construct into the host cells, can be used to
CC   modify the expression of G-CSF. These recombinant cells which express G-
CC   CSF are useful for in vitro production of the protein and gene therapy.
CC   Such cells may also be used in a delivery system for stimulating the
CC   proliferation and differentiation of haematopoietic progenitor cells, or
CC   for other conditions that can be treated with G-CSF, like chemotherapy-
CC   induced neutropenia, to treat patients undergoing bone marrow
CC   transplantation, chronic idiopathic and congenital neutropenic disorders
XX
SQ   Sequence 2834 BP; 670 A; 759 C; 817 G; 588 T; 0 U; 0 Other;
```

```
  Query Match             20.6%;  Score 412.2;  DB 3;  Length 2834;
  Best Local Similarity   65.9%;  Pred. No. 5e-89;
  Matches  805;  Conservative    0;  Mismatches  358;  Indels   58;  Gaps   12;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTC---------TAGCTAAAGGATTGTAAATGC 481
              || ||||||| |||||||||  ||||||||         || |||    ||       |
Db       1370 TCTGGTGGGGCCTTGGAGAATGTTTGTGTCGACACTCTGTATCTAGTTAATCTAGTGGGG 1429

Qy        482 ACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAA 541
              ||          || |||||||||| | |||||||||| |||||||||||  | |||| |
Db       1430 ACGTGGAGAACCTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCGCCCTGTCA 1489

Qy        542 AA--------------TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAA 587
              ||              |  ||||||||||||||||||||| |||| || ||||| || |||
Db       1490 AAACAGACCACTCGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAA 1549

Qy        588 AAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAG 647
              || | |||  |||||||||||||||||||| | | | ||||||  ||||| | ||| ||
Db       1550 AAGCAGGCTGCCCGAGCCAGCAGTGGCAACGCGCACAGGTCCCTATCCACAATATGGCAG 1609

Qy        648 CTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTA 707
              |||||||||||||||| ||  | |||||||||||| ||||||  | |||| |||||||||
Db       1610 CTTTGTTCTTTTGCTGTTTGCGATAAATCTTGCTACTGCTCGCTTTTTGGGTCCACACTG 1669

Qy        708 CCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AG 766
              | |||||||||||||||||||||  ||| ||||||  |||||| |||||||| ||   ||
Db       1670 CTTTTATGAGCTGTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCACTAAG 1729

Qy        767 ACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAAC 826
              ||||||| ||||| || |||||  || |||||| |   | |||||| ||||||||| ||||
Db       1730 ACCACGAGCCCACCGGGAGGAATGAACAACTCCGGCCGCGCTGCC-TTAAGAGCTATAAC 1788

Qy        827 ACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGA 886
              |||||| |||||| ||||||||||||||||||  || ||| |||||||| ||||||||||
Db       1789 ACTCACCGCGAAGGTCTGCAGCTTCACTCCT-CAGCCAGCGAGACCACGAACCCACCAGA 1847

Qy        887 AGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTT 946
              ||||||||||  | ||||| ||||| ||| ||||||||||||||||  ||||| | |
Db       1848 AGGAAGAAACTGCGAACACATCTGAACATCAGAAGGAACAAACTCCAGATGCACCA-CCT 1906

Qy        947 TCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCA 1006
              | |||||||||||||||| || |||||| | |||||| ||||||||||||  ||||||
Db       1907 TAAGAGCTGTAACACTCACTGCGAGGGTCCGCGGCTTCCTTCTTGAAGTCAGTGAGACCA 1966

Qy       1007 AGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTG 1066
              || || |||| |         |    | |      | |  |                 ||
Db       1967 AGCACTCACCAGTTTCGGACACAAGCCCAGGAGTTTGAGATCAGCCTGGGCAACATGATG 2026

Qy       1067 GCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATG 1126
                | |    |       |   |    ||   | ||          || |    |
Db       2027 AAATGCCCTCTCTGCAAAAAAAAAAAAAAATTACAAAAATTGGCGGAGCATGGTGGTCCGT 2086

Qy       1127 GTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATT 1186
              |  ||  |   ||  |  | |    | ||  |     |  |    ||  | | |||   |
Db       2087 GCCTGTGGTCCCAGCTACGCGGGAGGCTAAAGTGGGAGGATCGCTTGAGCCT-GGGAGGT 2145

Qy       1187 TCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTG 1246
                ||     |  |||||  |       |   |  | ||  ||  || | | | |
Db       2146 GAAGACTGCAGTGAGCTGTGATTGTACCACAGCCCTCTAGGCTGGGGGACAGACT----- 2200

Qy       1247 TAGACAGTAGTAGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAA 1306
                ||     || |   |  | | | | | ||  |    | |||  |
Db       2201 ----GAGACCCTGTTTCCCCTCCGCAAAAAAATTGACAAAAGTGTAATAAGAGGTGCCTG 2256

Qy       1307 ATATAGCCAGGTGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTG 1366
              |||| || ||| ||  |||||||||||||||||||||||||||||||| ||| ||| |.|
Db       2257 ATATGGCTAGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGG 2316

Qy       1367 TGGATCAC--GAGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTC 1424
               || ||||   |||||||| ||  |||||||||||||||||||| |||| ||| ||||
Db       2317 CGGGTCACCTAAGGTCAGGAGTGTGAGACCAGCCTGGCCAACATGGAGAAAGCCCATCTC 2376

Qy       1425 TACTAAAAATACAAAAATTAGCCT--------GGTGTGGTGGCACGCATCTGTAATCCCA 1476
              | ||||||||||||||| |   |          |  |||||||  |    ||||||||||
Db       2377 TTCTAAAAATACAAAATTAGCCGGCTGTGGGGGCAGTGGTGGAGCATGCCTGTAATCCCA 2436

Qy       1477 G-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGA 1535
              | |||||||||||||||||||| ||||| |||||||    |||||| | ||||||||||
Db       2437 GCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGCGGTTGCAGTGA 2496

Qy       1536 GCCAAGATCACACCACAGCACTCCATCC----------TGGGTGACAGAGCGAGACTCTG 1585
              ||| |||||    |||  |||||||| ||           ||     ||||| | ||||||
Db       2497 GCCGAGATCGTGCCATTGCACTCCACCCACTCCAGCCTGGGCAACAAGAGCCAAACTCTG 2556

Qy       1586 TCTCAAAAAAAAAAAAAAAAA 1606
              ||| ||||||||||||||||
Db       2557 TCTTAAAAAAAAAAAAAAAAA 2577


RESULT 33
AAZ29169
ID   AAZ29169 standard; DNA; 6235 BP.
XX
AC   AAZ29169;
XX
DT   21-FEB-2000  (first entry)
XX
DE   Human G-CSF genomic sequence upstream of transcription start site.
XX
KW   Granulocyte colony stimulating factor; G-CSF; genomic sequence;
KW   transcription start site; 5' non-coding sequence; DNA construct;
KW   targeting sequence; regulatory region; marker gene; selection;
KW   homologous recombination; gene therapy; delivery system; upstream;
KW   hematopoietic progenitor cell; chemotherapy-induced neutropenia;
KW   bone marrow transplantation; congenital neutropenic disorder;
KW   chronic idiopathic disorder; ds.
XX
OS   Homo sapiens.
XX
PN   WO9957291-A1.
XX
PD   11-NOV-1999.
XX
PF   05-MAY-1999;   99WO-US009924.
XX
PR   07-MAY-1998;   98US-0084649P.
XX
PA   (TRAN-) TRANSKARYOTIC THERAPIES INC.
XX
PI   Treco DA,  Heartlein MW,  Selden RF;
XX
DR   WPI; 2000-072235/06.
XX
PT   Novel genomic sequences used for treating human diseases and disorders.
XX
PS   Claim 1; Fig 5; 58pp; English.
XX
CC   The present DNA sequence is the human genomic sequence from a region
CC   upstream of the transcription start site of granulocyte colony-
CC   stimulating factor (G-CSF). This sequence represents nucleotides -6578 to
CC   -364 relative to the translation start site. G-CSF gene contains five
CC   exons and four introns. A DNA construct comprising a targeting sequence
CC   homologous to the 5'non-coding sequence of G-CSF, a transcriptional
CC   regulatory sequence that differs from the endogenous G-CSF gene, and a
CC   selectable marker gene for selection of recombinant cells is generated.
CC   It can be used by homologous recombination to modify the expression of
CC   endogenous G-CSF. These recombinant cells which express G-CSF are useful
CC   for in vitro production of the protein and gene therapy. Such cells may
CC   also be used in a delivery system for stimulating the proliferation and
CC   differentiation of hematopoietic progenitor cells, or for other
CC   conditions that can be treated with G-CSF, like chemotherapy-induced
CC   neutropenia, to treat patients undergoing bone marrow transplantation,
CC   chronic idiopathic and congenital neutropenic disorders
XX
SQ   Sequence 6235 BP; 1586 A; 1610 C; 1708 G; 1331 T; 0 U; 0 Other;
```

```
  Query Match             20.6%;  Score 412.2;  DB 3;  Length 6235;
  Best Local Similarity   65.9%;  Pred. No. 6.5e-89;
  Matches  805;  Conservative    0;  Mismatches  358;  Indels   58;  Gaps   12;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTC---------TAGCTAAAGGATTGTAAATGC 481
              || ||||||| |||||||||  ||||||||         || |||     ||       |
Db       3259 TCTGGTGGGGCCTTGGAGAATGTTTGTGTCGACACTCTGTATCTAGTTAATCTAGTGGGG 3318

Qy        482 ACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAA 541
              ||          || |||||||||||| | |||||||||| |||||||||||| | |||| |
Db       3319 ACGTGGAGAACCTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCGCCCTGTCA 3378

Qy        542 AA--------------TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAA 587
              ||              | ||||||||||||||||||||| |||| || ||||| || |||
Db       3379 AAACAGACCACTCGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAA 3438

Qy        588 AAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAG 647
              || | |||  |||||||||||||||||||| | | | ||||||  ||||||| | ||| ||
Db       3439 AAGCAGGCTGCCCGAGCCAGCAGTGGCAACGCGCACAGGTCCCTATCCACAATATGGCAG 3498

Qy        648 CTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTA 707
              |||||||||||||||| ||  | |||||||||||| ||||||  | |||| |||||||||
Db       3499 CTTTGTTCTTTTGCTGTTTGCGATAAATCTTGCTACTGCTCGCTTTTTGGGTCCACACTG 3558

Qy        708 CCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AG 766
              | |||||||||||||||||||||  ||| ||||||  |||||| |||||||| ||   ||
Db       3559 CTTTTATGAGCTGTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCACTAAG 3618

Qy        767 ACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAAC 826
              ||||||| ||||| || |||||  || |||||| |  | |||||| ||||||||| ||||
Db       3619 ACCACGAGCCCACCGGGAGGAATGAACAACTCCGGCCGCGCTGCC-TTAAGAGCTATAAC 3677

Qy        827 ACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGA 886
              |||||| |||||| ||||||||||||||||||  || ||| |||||||| ||||||||||
Db       3678 ACTCACCGCGAAGGTCTGCAGCTTCACTCCT-CAGCCAGCGAGACCACGAACCCACCAGA 3736

Qy        887 AGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTT 946
              ||||||||||  | ||||| ||||| ||| |||||||||||||||||  ||||| | |
Db       3737 AGGAAGAAACTGCGAACACATCTGAACATCAGAAGGAACAAACTCCAGATGCACCA-CCT 3795

Qy        947 TCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCA 1006
              | |||||||||||||||| || |||||| | |||||| ||||||||||||  ||||||
Db       3796 TAAGAGCTGTAACACTCACTGCGAGGGTCCGCGGCTTCCTTCTTGAAGTCAGTGAGACCA 3855

Qy       1007 AGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTG 1066
              || || |||| |         |    | |      | | |                   ||
Db       3856 AGCACTCACCAGTTTCGGACACAAGCCCAGGAGTTTGAGATCAGCCTGGGCAACATGATG 3915

Qy       1067 GCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATG 1126
              | |    |     |    |   ||   | ||          || |    |
Db       3916 AAATGCCCTCTCTGCAAAAAAAAAAAAAATTACAAAAATTGGCGGAGCATGGTGGTCCGT 3975

Qy       1127 GTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATT 1186
              |  ||  |   ||  |  | |    | ||  |     |  |    ||  | | |||   |
Db       3976 GCCTGTGGTCCCAGCTACGCGGGAGGCTAAAGTGGGAGGATCGCTTGAGCCT-GGGAGGT 4034

Qy       1187 TCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTG 1246
                ||     |  |||||  |       |    | ||  ||  || | | | |
Db       4035 GAAGACTGCAGTGAGCTGTGATTGTACCACAGCCCTCTAGGCTGGGGGACAGACT----- 4089

Qy       1247 TAGACAGTAGTAGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAA 1306
                ||    || |   |  |  | | |  | ||  |    | ||| |
Db       4090 ----GAGACCCTGTTTCCCCTCCGCAAAAAAATTGACAAAAGTGTAATAAGAGGTGCCTG 4145

Qy       1307 ATATAGCCAGGTGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTG 1366
              |||| || ||| ||  |||||||||||||||||||||||||||||||||| ||| ||| | |
Db       4146 ATATGGCTAGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGG 4205

Qy       1367 TGGATCAC--GAGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTC 1424
                || ||||   |||||||| ||  |||||||||||||||||||||| |||| ||| ||||
Db       4206 CGGGTCACCTAAGGTCAGGAGTGTGAGACCAGCCTGGCCAACATGGAGAAAGCCCATCTC 4265

Qy       1425 TACTAAAAATACAAAAATTAGCCT--------GGTGTGGTGGCACGCATCTGTAATCCCA 1476
              | |||||||||||||| |   |         |  |||||||  |    |||||||||||
Db       4266 TTCTAAAAATACAAAATTAGCCGGCTGTGGGGGCAGTGGTGGAGCATGCCTGTAATCCCA 4325

Qy       1477 G-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGA 1535
              | |||||||||||||||||||||| ||||| |||||||   |||||| | ||||||||||
Db       4326 GCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGCGGTTGCAGTGA 4385

Qy       1536 GCCAAGATCACACCACAGCACTCCATCC----------TGGGTGACAGAGCGAGACTCTG 1585
              ||| |||||   |||  |||||||| ||           ||     ||||| | ||||||
Db       4386 GCCGAGATCGTGCCATTGCACTCCACCCACTCCAGCCTGGGCAACAAGAGCCAAACTCTG 4445

Qy       1586 TCTCAAAAAAAAAAAAAAAAA 1606
              ||| ||||||||||||||||
Db       4446 TCTTAAAAAAAAAAAAAAAAA 4466
```

```
RESULT 34
AAZ29170
ID   AAZ29170 standard; DNA; 6679 BP.
XX
AC   AAZ29170;
XX
DT   21-FEB-2000  (first entry)
XX
DE   Partial sequence of human G-CSF upstream of transcription start site.
XX
KW   Granulocyte colony stimulating factor; G-CSF; partial genomic sequence;
KW   transcription start site; 5'non-coding sequence; DNA construct;
KW   targeting sequence; regulatory region; marker gene; selection;
KW   homologous recombination; gene therapy; delivery system; upstream;
KW   haematopoietic progenitor cell; chemotherapy-induced neutropenia;
KW   bone marrow transplantation; congenital neutropenic disorder;
KW   chronic idiopathic disorder; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .19
FT                   /*tag=  a
FT                   /note= "Sequence derived from the junction of genomic
FT                   insert and phage arm in the G-CSF/3 phage clone"
FT   TATA_signal     6550. .6555
FT                   /*tag=  b
FT                   /standard_name= "Hogness box"
FT   CDS             6618. .6679
FT                   /*tag=  c
FT                   /product= "N-terminal end of G-CSF"
FT                   /transl_except= (Pos:6648, AGPATQSPM)
FT                   /note= "no stop codon given"
XX
PN   WO9957291-A1.
XX
PD   11-NOV-1999.
XX
PF   05-MAY-1999;   99WO-US009924.
XX
PR   07-MAY-1998;   98US-0084649P.
XX
PA   (TRAN-) TRANSKARYOTIC THERAPIES INC.
XX
PI   Treco DA,  Heartlein MW,  Selden RF;
XX
DR   WPI; 2000-072235/06.
DR   P-PSDB; AAY15190.
XX
PT   Novel genomic sequences used for treating human diseases and disorders.
XX
PS   Disclosure; Page 48-49; 58pp; English.
XX
CC   The present DNA sequence is the human partial genomic sequence from a
CC   region upstream of the transcription start site of granulocyte colony-
CC   stimulating factor (G-CSF). A DNA construct comprising a targeting
CC   sequence homologous to the 5'non-coding sequence of G-CSF, a
```

```
CC   transcriptional regulatory sequence that differs from the endogenous G-
CC   CSF gene, and a selectable marker gene for selection of recombinant cells
CC   is generated, it can be used by homologous recombination to modify the
CC   expression of endogenous G-CSF. These recombinant cells which express G-
CC   CSF are useful for in vitro production of the protein and gene therapy.
CC   Such cells may also be used in a delivery system for stimulating the
CC   proliferation and differentiation of haematopoietic progenitor cells, or
CC   for other conditions that can be treated with G-CSF, like chemotherapy-
CC   induced neutropenia, to treat patients undergoing bone marrow
CC   transplantation, chronic idiopathic and congenital neutropenic disorders
XX
SQ   Sequence 6679 BP; 1684 A; 1739 C; 1830 G; 1426 T; 0 U; 0 Other;

  Query Match             20.6%;  Score 412.2;  DB 3;  Length 6679;
  Best Local Similarity   65.9%;  Pred. No. 6.7e-89;
  Matches  805;  Conservative    0;  Mismatches  358;  Indels   58;  Gaps   12;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTC---------TAGCTAAAGGATTGTAAATGC 481
              || ||||||| |||||||||  |||||||         || |||    ||       |
Db       3278 TCTGGTGGGGCCTTGGAGAATGTTTGTGTCGACACTCTGTATCTAGTTAATCTAGTGGGG 3337

Qy        482 ACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAA 541
              ||          || |||||||||| | |||||||||| |||||||||||| | |||| |
Db       3338 ACGTGGAGAACCTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCGCCCTGTCA 3397

Qy        542 AA--------------TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAA 587
              ||              | ||||||||||||||||||| |||| || ||||| || |||
Db       3398 AAACAGACCACTCGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAA 3457

Qy        588 AAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAG 647
              || | |||  |||||||||||||||||||| | | | ||||||  ||||||  | ||| ||
Db       3458 AAGCAGGCTGCCCGAGCCAGCAGTGGCAACGCGCACAGGTCCCTATCCACAATATGGCAG 3517

Qy        648 CTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTA 707
              |||||||||||||||| ||   | |||||||||||| ||||||   | |||| |||||||||
Db       3518 CTTTGTTCTTTTGCTGTTTGCGATAAATCTTGCTACTGCTCGCTTTTTGGGTCCACACTG 3577

Qy        708 CCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AG 766
              | |||||||||||||||||||   ||| ||||||  ||||||  |||||||| ||   ||
Db       3578 CTTTTATGAGCTGTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCACTAAG 3637

Qy        767 ACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAAC 826
              ||||||| ||||| || |||||  || |||||| |   | |||||| |||||||||  ||||
Db       3638 ACCACGAGCCCACCGGGAGGAATGAACAACTCCGGCCGCGCTGCC-TTAAGAGCTATAAC 3696

Qy        827 ACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGA 886
              ||||||  ||||||  ||||||||||||||||||   || |||  ||||||||  ||||||||||
Db       3697 ACTCACCGCGAAGGTCTGCAGCTTCACTCCT-CAGCCAGCGAGACCACGAACCCACCAGA 3755

Qy        887 AGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTT 946
              ||||||||||   | |||||  |||||  ||| ||||||||||||||||||  |||||  | |
Db       3756 AGGAAGAAACTGCGAACACATCTGAACATCAGAAGGAACAAACTCCAGATGCACCA-CCT 3814

Qy        947 TCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCA 1006
              | |||||||||||||||| || |||||| | ||||||  ||||||||||||||  ||||||
Db       3815 TAAGAGCTGTAACACTCACTGCGAGGGTCCGCGGCTTCCTTCTTGAAGTCAGTGAGACCA 3874

Qy       1007 AGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTG 1066
              || || ||||  |       |   |  |     | |  |                   ||
Db       3875 AGCACTCACCAGTTTCGGACACAAGCCCAGGAGTTTGAGATCAGCCTGGGCAACATGATG 3934

Qy       1067 GCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATG 1126
              | |  |      |  |  ||  | ||           || |    |
Db       3935 AAATGCCCTCTCTGCAAAAAAAAAAAAAAATTACAAAAATTGGCGGAGCATGGTGGTCCGT 3994

Qy       1127 GTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATT 1186
              |  ||  |  ||  |  | |    | ||  |     |  |    ||  | | |||   |
Db       3995 GCCTGTGGTCCCAGCTACGCGGGAGGCTAAAGTGGGAGGATCGCTTGAGCCT-GGGAGGT 4053

Qy       1187 TCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTG 1246
                ||    |  |||||  |       |   |  || ||  || | | | |
Db       4054 GAAGACTGCAGTGAGCTGTGATTGTACCACAGCCCTCTAGGCTGGGGGACAGACT----- 4108

Qy       1247 TAGACAGTAGTAGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAA 1306
                 ||     || |   |  | | | | | ||| |    | ||| |
Db       4109 ----GAGACCCTGTTTCCCCTCCGCAAAAAAATTGACAAAAGTGTAATAAGAGGTGCCTG 4164

Qy       1307 ATATAGCCAGGTGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTG 1366
              |||| || ||| ||  |||||||||||||||||||||||||||||||| ||| ||| | |
Db       4165 ATATGGCTAGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGG 4224

Qy       1367 TGGATCAC--GAGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTC 1424
                 || ||||   |||||||| ||  |||||||||||||||||||| |||| ||| ||||
Db       4225 CGGGTCACCTAAGGTCAGGAGTGTGAGACCAGCCTGGCCAACATGGAGAAAGCCCATCTC 4284

Qy       1425 TACTAAAAAATACAAAAATTAGCCT--------GGTGTGGTGGCACGCATCTGTAATCCCA 1476
              | |||||||||||||| |   |          |  |||||||  |    ||||||||||
Db       4285 TTCTAAAAATACAAAATTAGCCGGCTGTGGGGGCAGTGGTGGAGCATGCCTGTAATCCCA 4344

Qy       1477 G-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGA 1535
              | |||||||||||||||||||| ||||| |||||||   |||||| | ||||||||||
Db       4345 GCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGCGGTTGCAGTGA 4404

Qy       1536 GCCAAGATCACACCACAGCACTCCATCC----------TGGGTGACAGAGCGAGACTCTG 1585
              ||| |||||   |||  |||||||| ||           ||     ||||| | ||||||
Db       4405 GCCGAGATCGTGCCATTGCACTCCACCCACTCCAGCCTGGGCAACAAGAGCCAAACTCTG 4464

Qy       1586 TCTCAAAAAAAAAAAAAAAAA 1606
              ||| ||||||||||||||||
Db       4465 TCTTAAAAAAAAAAAAAAAAA 4485


RESULT 35
AAD17443
ID   AAD17443 standard; DNA; 6679 BP.
XX
AC   AAD17443;
XX
DT   10-DEC-2001  (first entry)
XX
DE   Human granulocyte colony stimulating factor (huGCSF) fragment #1.
XX
KW   Human; mutation; homologous recombination; target sequence; gene therapy;
KW   homologous recombination-enhancing agent; non-homologous end joining;
KW   therapeutic protein; granulocyte colony stimulating factor; huGCSF; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168882-A2.
XX
PD   20-SEP-2001.
XX
PF   13-MAR-2001; 2001WO-US007870.
XX
PR   14-MAR-2000; 2000US-00525160.
XX
PA   (TRAN-) TRANSKARYOTIC THERAPIES INC.
XX
PI   Ivanov E;
XX
DR   WPI; 2001-582459/65.
XX
PT   Complex or composition comprising a double stranded DNA sequence, a
PT   homologous recombination-enhancing agent, and agent inhibiting non-
PT   homologous end joining, for promoting alteration of a target sequence in
PT   a cell.
XX
PS   Disclosure; Page 77-79; 82pp; English.
XX
CC   The invention relates to a complex for promoting alteration of a target
CC   sequence in a cell, comprising a double stranded DNA sequence, a
CC   homologous recombination-enhancing agent and an agent inhibiting non-
CC   homologous end joining. The invention is used in gene therapy. The
CC   complex is useful for promoting an alteration at a selected site of a
CC   target sequence of a cell preferably of fungal, plant or animal origin,
CC   or of vertebrate origin which is a primary or secondary mammalian (human)
```

CC cell or an immortalised mammalian (human) cell, where target sequence
CC comprises a mutation preferably point mutation having less than 10 base
CC pairs which differ from wild-type sequence, (selected from cystic
CC fibrosis transmembrane regulator (CFTR) gene having mutation changes in
CC an amino acid encoded by codon 508, beta-globin gene having mutation
CC changes in an amino acid encoded by codon 6, Factor VIII gene having
CC mutation changes in an amino acid encoded by codon 2209 or 2229, Factor
CC IX gene, von Willebrand factor gene or xeroderma pigmentosa group G gene)
CC ; and the DNA sequence comprises a wild-type sequence which can correct
CC the mutation. The method further comprises introducing an agent which
CC inhibits a mismatch-repair protein (expression), which is from Msh2,
CC Msh6, Msh3, M1h1 and PMS2, or is an anti-mismatch-repair protein antibody
CC covalently linked to the DNA sequence, or to Rad52 protein or its
CC fragment. The complex is useful for altering expression of a protein
CC coding sequence of a gene in a cell. The method comprises introducing the
CC complex into the cell, where the DNA sequence comprises a regulatory
CC sequence, maintaining the cell under conditions which permit alteration
CC of a targetted genomic sequence to produce a homologously recombinant
CC cell and maintaining the homologously recombinant cell under conditions
CC which permit expression of the protein coding sequence of the gene under
CC control of the regulatory sequence. Homologously recombinant cell is
CC useful as a vehicle or delivery system for therapeutic proteins, such as
CC enzymes, hormones, cytokines, antigens, antibodies, clotting factors,
CC anti-sense RNA, regulatory proteins, transcription proteins, receptors,
CC structural proteins, novel (non-optimised) proteins and nucleic acid
CC products and engineered DNA and for supplying a therapeutic protein,
CC including erythropoietin, calcitonin, growth hormone, insulin and
CC insulinotropin. The present sequence is human granulocyte colony
CC stimulating factor (huGCSF) fragment used in the invention
XX
SQ Sequence 6679 BP; 1684 A; 1739 C; 1830 G; 1426 T; 0 U; 0 Other;

Query Match 20.6%; Score 412.2; DB 4; Length 6679;
Best Local Similarity 65.9%; Pred. No. 6.7e-89;
Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;

```
Qy       431 TCAGGTGGGGACTTGGAGAACTTTTGTGTC---------TAGCTAAAGGATTGTAAATGC 481
Db      3278 TCTGGTGGGGCCTTGGAGAATGTTTGTGTCGACACTCTGTATCTAGTTAATCTAGTGGGG 3337

Qy       482 ACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAA 541
Db      3338 ACGTGGAGAACCTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCGCCCTGTCA 3397

Qy       542 AA--------------TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAA 587
Db      3398 AAACAGACCACTCGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAA 3457

Qy       588 AAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAG 647
Db      3458 AAGCAGGCTGCCCGAGCCAGCAGTGGCAACGCGCACAGGTCCCTATCCACAATATGGCAG 3517

Qy       648 CTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTA 707
Db      3518 CTTTGTTCTTTTGCTGTTTGCGATAAATCTTGCTACTGCTCGCTTTTTGGGTCCACACTG 3577

Qy       708 CCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AG 766
Db      3578 CTTTTATGAGCTGTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCACTAAG 3637

Qy       767 ACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAAC 826
Db      3638 ACCACGAGCCCACCGGGAGGAATGAACAACTCCGGCCGCGCTGCC-TTAAGAGCTATAAC 3696

Qy       827 ACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGA 886
Db      3697 ACTCACCGCGAAGGTCTGCAGCTTCACTCCT-CAGCCAGCGAGACCACGAACCCACCAGA 3755

Qy       887 AGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTT 946
Db      3756 AGGAAGAAACTGCGAACACATCTGAACATCAGAAGGAACAAACTCCAGATGCACCA-CCT 3814

Qy       947 TCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCA 1006
Db      3815 TAAGAGCTGTAACACTCACTGCGAGGGTCCGCGGCTTCCTTCTTGAAGTCAGTGAGACCA 3874

Qy      1007 AGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTG 1066
Db      3875 AGCACTCACCAGTTTCGGACACAAGCCCAGGAGTTTGAGATCAGCCTGGGCAACATGATG 3934

Qy      1067 GCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATG 1126
Db      3935 AAATGCCCTCTCTGCAAAAAAAAAAAAAAATTACAAAAATTGGCGGAGCATGGTGGTCCGT 3994

Qy      1127 GTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATT 1186
Db      3995 GCCTGTGGTCCCAGCTACGCGGGAGGCTAAAGTGGGAGGATCGCTTGAGCCT-GGGAGGT 4053

Qy      1187 TCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTG 1246
Db      4054 GAAGACTGCAGTGAGCTGTGATTGTACCACAGCCCTCTAGGCTGGGGGACAGACT----- 4108

Qy      1247 TAGACAGTAGTAGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAA 1306
Db      4109 ----GAGACCCTGTTTCCCCTCCGCAAAAAAATTGACAAAAGTGTAATAAGAGGTGCCTG 4164

Qy      1307 ATATAGCCAGGTGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTG 1366
Db      4165 ATATGGCTAGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGG 4224

Qy      1367 TGGATCAC--GAGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTC 1424
Db      4225 CGGGTCACCTAAGGTCAGGAGTGTGAGACCAGCCTGGCCAACATGGAGAAAGCCCATCTC 4284

Qy      1425 TACTAAAAATACAAAAATTAGCCT--------GGTGTGGTGGCACGCATCTGTAATCCCA 1476
Db      4285 TTCTAAAAATACAAAATTAGCCGGCTGTGGGGGCAGTGGTGGAGCATGCCTGTAATCCCA 4344

Qy      1477 G-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGA 1535
Db      4345 GCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGCGGTTGCAGTGA 4404

Qy      1536 GCCAAGATCACACCACAGCACTCCATCC----------TGGGTGACAGAGCGAGACTCTG 1585
Db      4405 GCCGAGATCGTGCCATTGCACTCCACCCACTCCAGCCTGGGCAACAAGAGCCAAACTCTG 4464

Qy      1586 TCTCAAAAAAAAAAAAAAAAA 1606
Db      4465 TCTTAAAAAAAAAAAAAAAAA 4485
```

RESULT 36
AAD17447
ID AAD17447 standard; DNA; 6753 BP.
XX
AC AAD17447;
XX
DT 10-DEC-2001 (first entry)
XX
DE Human interferon alpha2 (huIFNalpha2) fragment #4.
XX
KW Human; mutation; homologous recombination; target sequence; gene therapy;
KW homologous recombination-enhancing agent; non-homologous end joining;
KW therapeutic protein; interferon alpha2; huIFNalpha2; ds.
XX
OS Homo sapiens.
XX
PN WO200168882-A2.
XX
PD 20-SEP-2001.
XX
PF 13-MAR-2001; 2001WO-US007870.
XX
PR 14-MAR-2000; 2000US-00525160.
XX
PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

```
XX
PI   Ivanov E;
XX
DR   WPI; 2001-582459/65.
XX
PT   Complex or composition comprising a double stranded DNA sequence, a
PT   homologous recombination-enhancing agent, and agent inhibiting non-
PT   homologous end joining, for promoting alteration of a target sequence in
PT   a cell.
XX
PS   Disclosure; Page 80-82; 82pp; English.
XX
CC   The invention relates to a complex for promoting alteration of a target
CC   sequence in a cell, comprising a double stranded DNA sequence, a
CC   homologous recombination-enhancing agent and an agent inhibiting non-
CC   homologous end joining. The invention is used in gene therapy. The
CC   complex is useful for promoting an alteration at a selected site of a
CC   target sequence of a cell preferably of fungal, plant or animal origin,
CC   or of vertebrate origin which is a primary or secondary mammalian (human)
CC   cell or an immortalised mammalian (human) cell, where target sequence
CC   comprises a mutation preferably point mutation having less than 10 base
CC   pairs which differ from wild-type sequence, (selected from cystic
CC   fibrosis transmembrane regulator (CFTR) gene having mutation changes in
CC   an amino acid encoded by codon 508, beta-globin gene having mutation
CC   changes in an amino acid encoded by codon 6, Factor VIII gene having
CC   mutation changes in an amino acid encoded by codon 2209 or 2229, Factor
CC   IX gene, von Willebrand factor gene or xeroderma pigmentosa group G gene)
CC   ; and the DNA sequence comprises a wild-type sequence which can correct
CC   the mutation. The method further comprises introducing an agent which
CC   inhibits a mismatch-repair protein (expression), which is from Msh2,
CC   Msh6, Msh3, M1h1 and PMS2, or is an anti-mismatch-repair protein antibody
CC   covalently linked to the DNA sequence, or to Rad52 protein or its
CC   fragment. The complex is useful for altering expression of a protein
CC   coding sequence of a gene in a cell. The method comprises introducing the
CC   complex into the cell, where the DNA sequence comprises a regulatory
CC   sequence, maintaining the cell under conditions which permit alteration
CC   of a targetted genomic sequence to produce a homologously recombinant
CC   cell and maintaining the homologously recombinant cell under conditions
CC   which permit expression of the protein coding sequence of the gene under
CC   control of the regulatory sequence. Homologously recombinant cell is
CC   useful as a vehicle or delivery system for therapeutic proteins, such as
CC   enzymes, hormones, cytokines, antigens, antibodies, clotting factors,
CC   anti-sense RNA, regulatory proteins, transcription proteins, receptors,
CC   structural proteins, novel (non-optimised) proteins and nucleic acid
CC   products and engineered DNA and for supplying a therapeutic protein,
CC   including erythropoietin, calcitonin, growth hormone, insulin and
CC   insulinotropin. The present sequence is human interferon alpha2
CC   (huIFNalpha2) fragment used in the invention
XX
SQ   Sequence 6753 BP; 2096 A; 1477 C; 1573 G; 1607 T; 0 U; 0 Other;

  Query Match           20.6%;  Score 412.2;  DB 4;  Length 6753;
  Best Local Similarity   65.9%;  Pred. No. 6.7e-89;
  Matches  805;  Conservative    0;  Mismatches  358;  Indels   58;  Gaps   12;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTC---------TAGCTAAAGGATTGTAAATGC 481
              || ||||||| |||||||||  ||||||||          ||  |||    ||       |
Db       2779 TCTGGTGGGGCCTTGGAGAATGTTTGTGTCGACACTCTGTATCTAGTTAATCTAGTGGGG 2838

Qy        482 ACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAA 541
              ||          || |||||||||| | |||||||||| |||||||||||| | |||| |
Db       2839 ACGTGGAGAACCTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCGCCCTGTCA 2898

Qy        542 AA--------------TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAA 587
              ||              | ||||||||||||||||||| |||| || ||||| || |||
Db       2899 AAACAGACCACTCGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAA 2958

Qy        588 AAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAG 647
              || | |||  |||||||||||||||||||| | | | ||||||  ||||||  | ||| ||
Db       2959 AAGCAGGCTGCCCGAGCCAGCAGTGGCAACGCGCACAGGTCCCTATCCACAATATGGCAG 3018

Qy        648 CTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTA 707
              |||||||||||||| ||  | |||||||||||| ||||||  | |||| |||||||||

Db       3019 CTTTGTTCTTTTGCTGTTTGCGATAAATCTTGCTACTGCTCGCTTTTTGGGTCCACACTG 3078

Qy        708 CCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AG 766
              | ||||||||||||||||||||||  ||| ||||||  ||||||  |||||||| ||   ||
Db       3079 CTTTTATGAGCTGTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCACTAAG 3138

Qy        767 ACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAAC 826
              ||||||| ||||| || |||||  || ||||||  |  | ||||||  |||||||||  ||||
Db       3139 ACCACGAGCCCACCGGGAGGAATGAACAACTCCGGCCGCGCTGCC-TTAAGAGCTATAAC 3197

Qy        827 ACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGA 886
              ||||||  ||||||  ||||||||||||||||||  || |||  ||||||||  ||||||||||
Db       3198 ACTCACCGCGAAGGTCTGCAGCTTCACTCCT-CAGCCAGCGAGACCACGAACCCACCAGA 3256

Qy        887 AGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTT 946
              ||||||||||  | |||||  |||||  |||  ||||||||||||||||||  |||||  | |
Db       3257 AGGAAGAAACTGCGAACACATCTGAACATCAGAAGGAACAAACTCCAGATGCACCA-CCT 3315

Qy        947 TCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCA 1006
              | ||||||||||||||||||  || ||||||  | ||||||  ||||||||||||||  ||||||
Db       3316 TAAGAGCTGTAACACTCACTGCGAGGGTCCGCGGCTTCCTTCTTGAAGTCAGTGAGACCA 3375

Qy       1007 AGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTG 1066
              || || |||| |      |   |  |  |  | |              ||
Db       3376 AGCACTCACCAGTTTCGGACACAAGCCCAGGAGTTTGAGATCAGCCTGGGCAACATGATG 3435

Qy       1067 GCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATG 1126
                | |    |     |   |  ||  | ||           || |    |
Db       3436 AAATGCCCTCTCTGCAAAAAAAAAAAAAAATTACAAAAATTGGCGGAGCATGGTGGTCCGT 3495

Qy       1127 GTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATT 1186
              |  ||  |   ||  |  |  |    | ||  |     |  |    || | | |||  |
Db       3496 GCCTGTGGTCCCAGCTACGCGGGAGGCTAAAGTGGGAGGATCGCTTGAGCCT-GGGAGGT 3554

Qy       1187 TCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTG 1246
                ||    |  |||||  |     |   |  | ||  ||  || | | | |
Db       3555 GAAGACTGCAGTGAGCTGTGATTGTACCACAGCCCTCTAGGCTGGGGGACAGACT----- 3609

Qy       1247 TAGACAGTAGTAGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAA 1306
                ||     || |   |  |  | | |  | ||  |    | |||
Db       3610 ----GAGACCCTGTTTCCCCTCCGCAAAAAAATTGACAAAAGTGTAATAAGAGGTGCCTG 3665

Qy       1307 ATATAGCCAGGTGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTG 1366
              |||| || ||| ||  |||||||||||||||||||||||||||||||||  ||| ||| | |
Db       3666 ATATGGCTAGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGG 3725

Qy       1367 TGGATCAC--GAGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTC 1424
                || ||||   |||||||| ||  |||||||||||||||||||| |||| ||| ||||
Db       3726 CGGGTCACCTAAGGTCAGGAGTGTGAGACCAGCCTGGCCAACATGGAGAAAGCCCATCTC 3785

Qy       1425 TACTAAAAATACAAAAATTAGCCT--------GGTGTGGTGGCACGCATCTGTAATCCCA 1476
              | |||||||||||||| |    |         |  |||||||  |    ||||||||||
Db       3786 TTCTAAAAATACAAAATTAGCCGGCTGTGGGGGCAGTGGTGGAGCATGCCTGTAATCCCA 3845

Qy       1477 G-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGA 1535
              | |||||||||||||||||||| ||||| |||||||    ||||||  | ||||||||||
Db       3846 GCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGCGGTTGCAGTGA 3905

Qy       1536 GCCAAGATCACACCACAGCACTCCATCC----------TGGGTGACAGAGCGAGACTCTG 1585
              |||  |||||   |||  ||||||||| ||          ||     |||||  | ||||||
Db       3906 GCCGAGATCGTGCCATTGCACTCCACCCACTCCAGCCTGGGCAACAAGAGCCAAACTCTG 3965

Qy       1586 TCTCAAAAAAAAAAAAAAAAA 1606
              |||  ||||||||||||||||
Db       3966 TCTTAAAAAAAAAAAAAAAAA 3986


RESULT 37
ADB62309
ID   ADB62309 standard; cDNA; 2942 BP.
XX
AC   ADB62309;
XX
```

```
DT   04-DEC-2003  (first entry)
XX
DE   Human cDNA encoding clone FCBBF30000010.
XX
KW   Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW   tissue regeneration; cell regeneration; membrane protein;
KW   signal transduction-related protein; transcription-related protein;
KW   osteoporosis; neurological disease; cancer; tumour.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             74. .643
FT                   /*tag=  a
FT                   /product= "Clone FCBBF30000010 protein"
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-450961/43.
DR   P-PSDB; ADB64279.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page; 222pp; English.
XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein to treat diseases. The
CC   sequence presented is a cDNA of the invention. Note: Some of the sequence
CC   data for this patent is not represented in the printed specification, but
CC   is based on sequence information supplied by the European Patent Office.
XX
SQ   Sequence 2942 BP; 785 A; 743 C; 728 G; 686 T; 0 U; 0 Other;

  Query Match            20.0%;  Score 401.2;  DB 10;  Length 2942;
  Best Local Similarity  83.7%;  Pred. No. 2.3e-86;
  Matches  497;  Conservative    0;  Mismatches   78;  Indels   19;  Gaps    3;

Qy        453 TTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGA 512
              | ||||||||||| |||| ||||||||||||||||||||||||| ||

Db       2116 TCTGTGTCTAGCTGAAGGTTTGTAAATGCACCAATCAGCACTCCGTA---------- 2162

Qy        513 TTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGT 572
              ||| | ||  ||||| || ||||||||||||||||||||||||||| ||| |||

Db       2163 ----AAACGGACTGATCAGTGCTTTGTAAAATGGACCAATCAGCAGGATGTAGGCAGGGC 2218

Qy        573 CAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCT 632
              |||||||||||  |||| |||| ||||| ||||| ||| |||||||  || | ||||| |

Db       2219 CAAATAAGGGAAAAAAAGCTGG-CACCCAAGCCAACAGGGGCAACCTGCTTGAGTCCCTT 2277

Qy        633 TCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTC 692
              |||||| |||||||||||||||| ||  |||||||  |||||||||||||||||||| ||

Db       2278 TCCACATTGTGGAAGCTTTGTTCCTTCACTCTTCATGATAAATCTTGCTGCTGCTCACTC 2337

Qy        693 TTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATT 752
              |||| |||| ||||||||||||||||||||||||||||||  ||||||||||||||||||

Db       2338 TTTGGGTCCGCACTACCTTTATGAGCTGTAACACTCACCACGAGGGTCTGTGGCTTCATT 2397

Qy        753 CCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCC 811
              ||| |||| | | |||||||||||||||| || ||||||||| ||||||  |      |  ||

Db       2398 CCTTAAGTTAGCAAGACCACGAACCCACAGGGAGGAACAAACAACTCCGGGCACACCACC 2457

Qy        812 TTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGAC 871
              |||||||||| |||||||||||||||| ||||  |||||||||||| ||||||  | ||

Db       2458 TTTAAGAGCTATAACACTCACTGCGAAGGTCTGTGGCTTCACTCCTGTAGTCAGCAAAAC 2517

Qy        872 CACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTC 931
              |||||||||||  |||||||||||||  |||||| || || ||||||||||||||||||

Db       2518 CACAAACCCACCGGAAGGAAGAAACTCCAGACACATCTAAACATCTGAAGGAACAAACTC 2577

Qy        932 CAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTG 991
              |||||||||||||| ||| ||||||||||||||  |||||| ||||||||||||||||

Db       2578 TGGACACACCATCTTTAAGAACTGTAACACTCACCGTGAGGGTCCGTGGCTTCATTCTTG 2637

Qy        992 AAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
              |||||||  |||||||||||||||||||||||||||| ||| |||||| ||| |

Db       2638 AAGTCAGGGAGACCAAGAACCCACCGGAAGGAACAAACTCCGGACACACTAGCA 2691


RESULT 38
ABZ74190/c
ID   ABZ74190 standard; DNA; 27666 BP.
XX
AC   ABZ74190;
XX
DT   12-MAY-2003  (first entry)
XX
DE   Secreted protein gene 215 genomic fragment HNGAM58, SEQ ID NO:1337.
XX
KW   Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KW   autoimmune disorder; inflammation; angiogenic diseases; AIDS;
KW   acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KW   drug screening; chromosome identification; chromosome mapping;
KW   cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
KW   antianaemic; vulnerary; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277013-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-US009370.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-040578/03.
```

```
XX
PT   New human secreted proteins and nucleic acids, useful for detecting or
PT   treating cancer or other hyperproliferative disorders, autoimmune
PT   disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX
PS   Disclosure; Page 1963-1970; 2474pp; English.
XX
CC   ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
CC   protein genes, and ABP00947-ABP01363 represent the proteins they encode.
CC   ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC   invention also encompasses antibodies specific for the secreted proteins,
CC   the use of the secreted proteins in drug screening and recombinant
CC   vectors and host cells comprising a nucleic acid of the invention. The
CC   secreted proteins are thought to be involved in biological activities
CC   associated with cellular signalling, cellular differentiation, cell
CC   migration, prohormone activation and neurotransmitter activity. The
CC   secreted proteins, nucleic acids encoding them, antibodies or antibody
CC   fragments specific for the secreted proteins, and modulators of protein
CC   activity are useful for diagnosing or treating cancers or other
CC   hyperproliferative disorders. Additionally, the secreted proteins and
CC   their nucleic acids may also be used in the treatment of autoimmune
CC   disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC   (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC   wound healing. Nucleic acids of the invention may be used for chromosome
CC   identification, chromosome mapping, in gene therapy, for identifying
CC   individuals from minute biological samples, as hybridisation probes, and
CC   as molecular weight markers. The present sequence represents a human
CC   secreted protein genomic fragment referred to in the disclosure of the
CC   invention
XX
SQ   Sequence 27666 BP; 7795 A; 6389 C; 6455 G; 7027 T; 0 U; 0 Other;

  Query Match             20.0%;  Score 400.8;  DB 8;  Length 27666;
  Best Local Similarity   80.7%;  Pred. No. 6.1e-86;
  Matches  499;  Conservative    0;  Mismatches  102;  Indels   17;  Gaps    2;

Qy        461 TAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAAT 520
              |||||||||| ||||||| ||||||||||||||||||||                ||||
Db      22694 TAGCTAAAGGTTTGTAAAGGCACCAATCAGCACTCTGTAA----------------AAAT 22651

Qy        521 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAG 580
              | |||||||||| ||||||||||||||||||||||||||||||| |||| ||||||||
Db      22650 GGACCAATCAGCGCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAG 22591

Qy        581 GGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACT 640
              ||| ||||| |||||||||| | |||||| | |||||||| || |||| | ||| ||  ||
Db      22590 GGAATAAAAGCTGGCCACCTGCGCCAGCGGGGGCAACCCGCTGGGGTACTCTTACATGCT 22531

Qy        641 GTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTC 700
              ||||| |||||||||||||||||||   || |||||||||| |||||||| |||| |||
Db      22530 GTGGAGGCTTTGTTCTTTTGCTCTTTGCAGTAAATCTTGCTGTTGCTCATTTTTTGGGTC 22471

Qy        701 CACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGT 760
                 |||||||||||||||||||||||||  | || ||||||  ||||||| |||| ||||
Db      22470 TGTACTACCTTTATGAGCTGTAACACTCAGCGTGAAGGTCTGCAGCTTCACTCCTCAAGT 22411

Qy        761 CAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGC 820
              || ||   |    | || |  | ||||||||| ||||||| | ||||  ||||||||||
Db      22410 CAGCAATACCACGAACCCCCCGGAGGAACAAACAACTCCCAACGTGCCACCTTTAAGAGC 22351

Qy        821 TGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCC 880
              |||||||||||||||||| |||||  ||| |||||||||||||||||||| |||||
Db      22350 TGTAACACTCACTGCGAAGGTCTGCGACTTTACTCCTGAAGTCAGTGAGACCACGAACCC 22291

Qy        881 ACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACAC 940
              |||||||||||||||||  |  ||   |||| ||||||||||||||| ||||||||||
Db      22290 ACCAGAAGGAAGAAACTCCAGGTACCTGTGAACATCTGAAGGAACAAACGCCAGACACAC 22231

Qy        941 CATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCA 1000
              || |||| ||| ||| ||||||||| || ||||||    ||||||||||||||||||||
Db      22230 CAGCTTTAAGAACTGCAACACTCACTGCGAGGGTC-CCAGCTTCATTCTTGAAGTCAGCG 22172

Qy       1001 AGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGA 1060
              |||||| ||||||||| |||||||| ||||||||||||||||  |  || |  |  |  |
Db      22171 AGACCAGGAACCCACCAGAAGGAACCAATTCCAGACACAGTACAATCTCAGCTTACTGCA 22112

Qy       1061 TCTGTGGCTTCCAGGGTT 1078
              |    || ||| |||||
Db      22111 ACCTCTGCCTCCTGGGTT 22094


RESULT 39
ADA98719/c
ID   ADA98719 standard; DNA; 27666 BP.
XX
AC   ADA98719;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human secreted protein-related DNA sequence #312.
XX
KW   human; secreted protein; cardiovascular disorder; arrhythmia;
KW   atherosclerosis; stroke; endocarditis; congestive heart failure;
KW   rheumatic heart disease; cardiomyopathy; hemorrhoids; varicose veins;
KW   migraine; thrombosis; neural disorder; immune system disorder;
KW   muscular disorder; reproductive disorder; gastrointestinal disorder;
KW   pulmonary disorder; renal disorder; proliferative disorder; cancer; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003004623-A2.
XX
PD   16-JAN-2003.
XX
PF   26-MAR-2002; 2002WO-US009922.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-247946/24.
XX
PT   New human secreted polypeptide and nucleic acid molecules, useful for
PT   diagnosing, preventing, prognosticating or treating cardiovascular
PT   disorders (e.g. arrhythmia, atherosclerosis, cardiomyopathy, or
PT   thrombosis).
XX
PS   Disclosure; SEQ ID NO 828; 1572pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of human
CC   secreted proteins. The DNA and protein sequences of the invention are
CC   useful in the treatment of cardiovascular disorders, such as: arrhythmia,
CC   atherosclerosis, stroke, endocarditis, congestive heart failure,
CC   rheumatic heart disease, cardiomyopathy, hemorrhoids, varicose veins,
CC   migraine, or thrombosis. The DNA and protein sequences may also be used
CC   for treating or preventing: neural disorders, immune system disorders,
CC   muscular disorders, reproductive disorders, gastrointestinal disorders,
CC   pulmonary disorders, renal disorders, proliferative disorders and/or
CC   cancerous diseases. The present DNA sequence is used in the
CC   exemplification of the invention. NOTE: The present sequence is shown on
CC   the WIPO website.
XX
SQ   Sequence 27666 BP; 7795 A; 6389 C; 6455 G; 7027 T; 0 U; 0 Other;

  Query Match             20.0%;  Score 400.8;  DB 8;  Length 27666;
  Best Local Similarity   80.7%;  Pred. No. 6.1e-86;
  Matches  499;  Conservative    0;  Mismatches  102;  Indels   17;  Gaps    2;

Qy        461 TAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAAT 520
              |||||||||| ||||||| ||||||||||||||||||||                ||||
Db      22694 TAGCTAAAGGTTTGTAAAGGCACCAATCAGCACTCTGTAA----------------AAAT 22651
```

```
Qy        521 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAG 580
              | |||||||||| |||||||||||||||||||||||||||||||||| |||| ||||||||
Db      22650 GGACCAATCAGCGCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAG 22591

Qy        581 GGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACT 640
              ||| ||||| ||||||||| | |||||| | |||||||| || |||| | ||| ||  ||
Db      22590 GGAATAAAAGCTGGCCACCTGCGCCAGCGGGGGCAACCCGCTGGGGTACTCTTACATGCT 22531

Qy        641 GTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTC 700
              ||||| |||||||||||||||||  || |||||||||||| |||||||| |||| |||
Db      22530 GTGGAGGCTTTGTTCTTTTGCTCTTTGCAGTAAATCTTGCTGTTGCTCATTTTTTGGGTC 22471

Qy        701 CACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGT 760
              |||||||||||||||||||||||||  | || ||||||  |||||| |||| ||||
Db      22470 TGTACTACCTTTATGAGCTGTAACACTCAGCGTGAAGGTCTGCAGCTTCACTCCTCAAGT 22411

Qy        761 CAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGC 820
              || ||   |    | || |  | ||||||||| ||||||| | ||||  ||||||||||
Db      22410 CAGCAATACCACGAACCCCCCGGAGGAACAAACAACTCCCAACGTGCCACCTTTAAGAGC 22351

Qy        821 TGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCC 880
              |||||||||||||||||| |||||  ||| |||||||||||||||||||||| |||||
Db      22350 TGTAACACTCACTGCGAAGGTCTGCGACTTTACTCCTGAAGTCAGTGAGACCACGAACCC 22291

Qy        881 ACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACAC 940
              |||||||||||||||||  |  ||   |||| |||||||||||||||| ||||||||||
Db      22290 ACCAGAAGGAAGAAACTCCAGGTACCTGTGAACATCTGAAGGAACAAACGCCAGACACAC 22231

Qy        941 CATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCA 1000
              || |||| ||| ||| |||||||||| || ||||||    ||||||||||||||||||||
Db      22230 CAGCTTTAAGAACTGCAACACTCACTGCGAGGGTC-CCAGCTTCATTCTTGAAGTCAGCG 22172

Qy       1001 AGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGA 1060
              |||||| ||||||||| |||||||| ||||||||||||||  |  || |  |  |  |
Db      22171 AGACCAGGAACCCACCAGAAGGAACCAATTCCAGACACAGTACAATCTCAGCTTACTGCA 22112

Qy       1061 TCTGTGGCTTCCAGGGTT 1078
              |    || ||| |||||
Db      22111 ACCTCTGCCTCCTGGGTT 22094
```

RESULT 40
ABZ67756/c
ID ABZ67756 standard; DNA; 27666 BP.
XX
AC ABZ67756;
XX
DT 26-MAR-2003 (first entry)
XX
DE Human secreted protein encoding genomic DNA SEQ ID NO 1279.
XX
KW Human; secreted protein; nootropic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW vulnerary; antibacterial; antiparkinsonian; antisickling; antianaemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277186-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-US009188.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040583/03.
XX
PT New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
PS Disclosure; Page 1946-1953; 2423pp; English.
XX
CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 27666 BP; 7795 A; 6389 C; 6455 G; 7027 T; 0 U; 0 Other;

Query Match 20.0%; Score 400.8; DB 10; Length 27666;
Best Local Similarity 80.7%; Pred. No. 6.1e-86;
Matches 499; Conservative 0; Mismatches 102; Indels 17; Gaps 2;

```
Qy        461 TAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAAT 520
              |||||||||| ||||||| ||||||||||||||||||                ||||
Db      22694 TAGCTAAAGGTTTGTAAAGGCACCAATCAGCACTCTGTAA----------------AAAT 22651

Qy        521 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAG 580
              | |||||||||| |||||||||||||||||||||||||||||||||| |||| ||||||||
Db      22650 GGACCAATCAGCGCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAG 22591

Qy        581 GGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACT 640
              ||| ||||| ||||||||| | |||||| | |||||||| || |||| | ||| ||  ||
Db      22590 GGAATAAAAGCTGGCCACCTGCGCCAGCGGGGGCAACCCGCTGGGGTACTCTTACATGCT 22531

Qy        641 GTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTC 700
              ||||| |||||||||||||||||  || |||||||||||| |||||||| |||| |||
Db      22530 GTGGAGGCTTTGTTCTTTTGCTCTTTGCAGTAAATCTTGCTGTTGCTCATTTTTTGGGTC 22471

Qy        701 CACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGT 760
              |||||||||||||||||||||||||  | || ||||||  |||||| |||| ||||
Db      22470 TGTACTACCTTTATGAGCTGTAACACTCAGCGTGAAGGTCTGCAGCTTCACTCCTCAAGT 22411

Qy        761 CAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGC 820
              || ||   |    | || |  | ||||||||| ||||||| | ||||  ||||||||||
Db      22410 CAGCAATACCACGAACCCCCCGGAGGAACAAACAACTCCCAACGTGCCACCTTTAAGAGC 22351

Qy        821 TGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCC 880
              |||||||||||||||||| |||||  ||| |||||||||||||||||||||| |||||
Db      22350 TGTAACACTCACTGCGAAGGTCTGCGACTTTACTCCTGAAGTCAGTGAGACCACGAACCC 22291

Qy        881 ACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACAC 940
              |||||||||||||||||  |  ||   |||| |||||||||||||||| ||||||||||
Db      22290 ACCAGAAGGAAGAAACTCCAGGTACCTGTGAACATCTGAAGGAACAAACGCCAGACACAC 22231

Qy        941 CATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCA 1000
              || |||| ||| ||| |||||||||| || ||||||    ||||||||||||||||||||
Db      22230 CAGCTTTAAGAACTGCAACACTCACTGCGAGGGTC-CCAGCTTCATTCTTGAAGTCAGCG 22172

Qy       1001 AGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGA 1060
              |||||| ||||||||| |||||||| ||||||||||||||||  |  || |  |  |  |
Db      22171 AGACCAGGAACCCACCAGAAGGAACCAATTCCAGACACAGTACAATCTCAGCTTACTGCA 22112
```

```
Qy       1061 TCTGTGGCTTCCAGGGTT 1078
                |    || ||| |||||
Db      22111 ACCTCTGCCTCCTGGGTT 22094
```

RESULT 41
AAL06082
ID AAL06082 standard; DNA; 32212 BP.
XX
AC AAL06082;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 8770.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.

```
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is
PT   used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 8770; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention
XX
SQ   Sequence 32212 BP; 8041 A; 7564 C; 7490 G; 9117 T; 0 U; 0 Other;

  Query Match             20.0%;  Score 400.8;  DB 4;  Length 32212;
  Best Local Similarity   80.7%;  Pred. No. 6.4e-86;
  Matches  499;  Conservative    0;  Mismatches  102;  Indels   17;  Gaps    2;

Qy        461 TAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAAT 520
              |||||||||| ||||||| ||||||||||||||||||                ||||
Db       4974 TAGCTAAAGGTTTGTAAAGGCACCAATCAGCACTCTGTAA----------------AAAT 5017

Qy        521 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAG 580
              | |||||||||| ||||||||||||||||||||||||||||||||||| |||| |||||||||
Db       5018 GGACCAATCAGCGCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAG 5077

Qy        581 GGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACT 640
              ||| ||||| ||||||||| | |||||| | |||||||| || |||| | ||| ||  ||
Db       5078 GGAATAAAAGCTGGCCACCTGCGCCAGCGGGGGCAACCCGCTGGGGTACTCTTACATGCT 5137

Qy        641 GTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTC 700
              ||||| |||||||||||||||||||  || |||||||||||| |||||||| |||| |||
Db       5138 GTGGAGGCTTTGTTCTTTTGCTCTTTGCAGTAAATCTTGCTGTTGCTCATTTTTTGGGTC 5197

Qy        701 CACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGT 760
              ||||||||||||||||||||||||  | || ||||||  |||||| |||| ||||
Db       5198 TGTACTACCTTTATGAGCTGTAACACTCAGCGTGAAGGTCTGCAGCTTCACTCCTCAAGT 5257

Qy        761 CAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGC 820
              || ||   |    | || |  | ||||||||| ||||||| | ||||  ||||||||||
Db       5258 CAGCAATACCACGAACCCCCCGGAGGAACAAACAACTCCCAACGTGCCACCTTTAAGAGC 5317

Qy        821 TGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCC 880
              |||||||||||||||||| |||||  ||| ||||||||||||||||||||| |||||
Db       5318 TGTAACACTCACTGCGAAGGTCTGCGACTTTACTCCTGAAGTCAGTGAGACCACGAACCC 5377

Qy        881 ACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACAC 940
              |||||||||||||||| |  ||   |||| |||||||||||||| ||||||||||
Db       5378 ACCAGAAGGAAGAAACTCCAGGTACCTGTGAACATCTGAAGGAACAAACGCCAGACACAC 5437

Qy        941 CATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCA 1000
              || |||| ||| ||| ||||||||| || ||||||    ||||||||||||||||||||
Db       5438 CAGCTTTAAGAACTGCAACACTCACTGCGAGGGTC-CCAGCTTCATTCTTGAAGTCAGCG 5496

Qy       1001 AGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGA 1060
              |||||| ||||||||| |||||||| |||||||||||||||  |  || |  |  | |
Db       5497 AGACCAGGAACCCACCAGAAGGAACCAATTCCAGACACAGTACAATCTCAGCTTACTGCA 5556

Qy       1061 TCTGTGGCTTCCAGGGTT 1078
                |    || ||| |||||
Db       5557 ACCTCTGCCTCCTGGGTT 5574


RESULT 42
ABL98647
ID   ABL98647 standard; DNA; 32212 BP.
XX
AC   ABL98647;
XX
DT   21-JUN-2002  (first entry)
XX
DE   Human testicular antigen encoding DNA fragment SEQ ID NO: 3299.
XX
KW   Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW   reproductive system disorder; urinary system disorder; gene therapy;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155317-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001329.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
```

PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483232/52.
XX
PT Nucleic acids encoding 973 human testicular antigen polypeptides, useful
PT for preventing, diagnosing and/or treating testicular cancer.
XX
PS Disclosure; SEQ ID NO 3299; 766pp; English.
XX
CC The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a DNA encoding a
CC protein fragment of the invention
XX
SQ Sequence 32212 BP; 8041 A; 7564 C; 7490 G; 9117 T; 0 U; 0 Other;

Query Match 20.0%; Score 400.8; DB 4; Length 32212;
Best Local Similarity 80.7%; Pred. No. 6.4e-86;
Matches 499; Conservative 0; Mismatches 102; Indels 17; Gaps 2;

```
Qy       461 TAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAAT 520
            |||||||||| ||||||| ||||||||||||||||||||                ||||
Db      4974 TAGCTAAAGGTTTGTAAAGGCACCAATCAGCACTCTGTAA----------------AAAT 5017

Qy       521 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAG 580
             | |||||||||| ||||||||||||||||||||||||||||||||| |||| ||||||||
Db      5018 GGACCAATCAGCGCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAG 5077

Qy       581 GGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACT 640
             ||| ||||| ||||||||| | |||||| | |||||||| || |||| | ||| ||  ||
Db      5078 GGAATAAAAGCTGGCCACCTGCGCCAGCGGGGGCAACCCGCTGGGGTACTCTTACATGCT 5137

Qy       641 GTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTC 700
             ||||| |||||||||||||||||  || |||||||||||| |||||||| |||| |||
Db      5138 GTGGAGGCTTTGTTCTTTTGCTCTTTGCAGTAAATCTTGCTGTTGCTCATTTTTTGGGTC 5197

Qy       701 CACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGT 760
               |||||||||||||||||||||||||  | || ||||||  |||||| |||| ||||
Db      5198 TGTACTACCTTTATGAGCTGTAACACTCAGCGTGAAGGTCTGCAGCTTCACTCCTCAAGT 5257

Qy       761 CAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGC 820
```

```
Db       5258 CAGCAATACCACGAACCCCCCGGAGGAACAAACAACTCCCAACGTGCCACCTTTAAGAGC 5317

Qy        821 TGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCC 880

Db       5318 TGTAACACTCACTGCGAAGGTCTGCGACTTTACTCCTGAAGTCAGTGAGACCACGAACCC 5377

Qy        881 ACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACAC 940

Db       5378 ACCAGAAGGAAGAAACTCCAGGTACCTGTGAACATCTGAAGGAACAAACGCCAGACACAC 5437

Qy        941 CATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCA 1000

Db       5438 CAGCTTTAAGAACTGCAACACTCACTGCGAGGGTC-CCAGCTTCATTCTTGAAGTCAGCG 5496

Qy       1001 AGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGA 1060

Db       5497 AGACCAGGAACCCACCAGAAGGAACCAATTCCAGACACAGTACAATCTCAGCTTACTGCA 5556

Qy       1061 TCTGTGGCTTCCAGGGTT 1078

Db       5557 ACCTCTGCCTCCTGGGTT 5574
```

RESULT 43
ABZ74191/c
ID ABZ74191 standard; DNA; 41206 BP.
XX
AC ABZ74191;
XX
DT 12-MAY-2003 (first entry)
XX
DE Secreted protein gene 215 genomic fragment HNGAM58, SEQ ID NO:1338.
XX
KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KW drug screening; chromosome identification; chromosome mapping;
KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
KW antianaemic; vulnerary; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277013-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-US009370.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040578/03.
XX
PT New human secreted proteins and nucleic acids, useful for detecting or
PT treating cancer or other hyperproliferative disorders, autoimmune
PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX
PS Disclosure; Page 1970-1980; 2474pp; English.
XX
CC ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins are thought to be involved in biological activities
CC associated with cellular signalling, cellular differentiation, cell
CC migration, prohormone activation and neurotransmitter activity. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein genomic fragment referred to in the disclosure of the
CC invention
XX
SQ Sequence 41206 BP; 11685 A; 9614 C; 9725 G; 10182 T; 0 U; 0 Other;

```
  Query Match             20.0%;  Score 400.8;  DB 8;  Length 41206;
  Best Local Similarity   80.7%;  Pred. No. 6.9e-86;
  Matches  499;  Conservative    0;  Mismatches  102;  Indels   17;  Gaps    2;

Qy         461 TAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAAT 520

Db       36233 TAGCTAAAGGTTTGTAAAGGCACCAATCAGCACTCTGTAA----------------AAAT 36190

Qy         521 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAG 580

Db       36189 GGACCAATCAGCGCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAG 36130

Qy         581 GGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACT 640

Db       36129 GGAATAAAAGCTGGCCACCTGCGCCAGCGGGGGCAACCCGCTGGGGTACTCTTACATGCT 36070

Qy         641 GTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTC 700

Db       36069 GTGGAGGCTTTGTTCTTTTGCTCTTTGCAGTAAATCTTGCTGTTGCTCATTTTTTGGGTC 36010

Qy         701 CACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGT 760

Db       36009 TGTACTACCTTTATGAGCTGTAACACTCAGCGTGAAGGTCTGCAGCTTCACTCCTCAAGT 35950

Qy         761 CAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGC 820

Db       35949 CAGCAATACCACGAACCCCCCGGAGGAACAAACAACTCCCAACGTGCCACCTTTAAGAGC 35890

Qy         821 TGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCC 880

Db       35889 TGTAACACTCACTGCGAAGGTCTGCGACTTTACTCCTGAAGTCAGTGAGACCACGAACCC 35830

Qy         881 ACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACAC 940

Db       35829 ACCAGAAGGAAGAAACTCCAGGTACCTGTGAACATCTGAAGGAACAAACGCCAGACACAC 35770

Qy         941 CATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCA 1000

Db       35769 CAGCTTTAAGAACTGCAACACTCACTGCGAGGGTC-CCAGCTTCATTCTTGAAGTCAGCG 35711

Qy        1001 AGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGA 1060

Db       35710 AGACCAGGAACCCACCAGAAGGAACCAATTCCAGACACAGTACAATCTCAGCTTACTGCA 35651

Qy        1061 TCTGTGGCTTCCAGGGTT 1078

Db       35650 ACCTCTGCCTCCTGGGTT 35633
```

RESULT 44
ADA98720/c
ID ADA98720 standard; DNA; 41206 BP.
XX
AC ADA98720;
XX
DT 20-NOV-2003 (first entry)

```
XX
DE   Human secreted protein-related DNA sequence #313.
XX
KW   human; secreted protein; cardiovascular disorder; arrhythmia;
KW   atherosclerosis; stroke; endocarditis; congestive heart failure;
KW   rheumatic heart disease; cardiomyopathy; hemorrhoids; varicose veins;
KW   migraine; thrombosis; neural disorder; immune system disorder;
KW   muscular disorder; reproductive disorder; gastrointestinal disorder;
KW   pulmonary disorder; renal disorder; proliferative disorder; cancer; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003004623-A2.
XX
PD   16-JAN-2003.
XX
PF   26-MAR-2002; 2002WO-US009922.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-247946/24.
XX
PT   New human secreted polypeptide and nucleic acid molecules, useful for
PT   diagnosing, preventing, prognosticating or treating cardiovascular
PT   disorders (e.g. arrhythmia, atherosclerosis, cardiomyopathy, or
PT   thrombosis).
XX
PS   Disclosure; SEQ ID NO 829; 1572pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of human
CC   secreted proteins. The DNA and protein sequences of the invention are
CC   useful in the treatment of cardiovascular disorders, such as: arrhythmia,
CC   atherosclerosis, stroke, endocarditis, congestive heart failure,
CC   rheumatic heart disease, cardiomyopathy, hemorrhoids, varicose veins,
CC   migraine, or thrombosis. The DNA and protein sequences may also be used
CC   for treating or preventing: neural disorders, immune system disorders,
CC   muscular disorders, reproductive disorders, gastrointestinal disorders,
CC   pulmonary disorders, renal disorders, proliferative disorders and/or
CC   cancerous diseases. The present DNA sequence is used in the
CC   exemplification of the invention. NOTE: The present sequence is shown on
CC   the WIPO website.
XX
SQ   Sequence 41206 BP; 11685 A; 9614 C; 9725 G; 10182 T; 0 U; 0 Other;

  Query Match             20.0%;  Score 400.8;  DB 8;  Length 41206;
  Best Local Similarity   80.7%;  Pred. No. 6.9e-86;
  Matches  499;  Conservative    0;  Mismatches  102;  Indels   17;  Gaps    2;

Qy        461 TAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAAT 520
              |||||||||| ||||||| ||||||||||||||||||||                 ||||
Db      36233 TAGCTAAAGGTTTGTAAAGGCACCAATCAGCACTCTGTAA----------------AAAT 36190

Qy        521 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAG 580
              | |||||||||| ||||||||||||||||||||||||||||||| |||| ||||||||
Db      36189 GGACCAATCAGCGCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAG 36130

Qy        581 GGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACT 640
              ||| ||||| ||||||||| | |||||| | |||||||| || |||| | ||| ||  ||
Db      36129 GGAATAAAAGCTGGCCACCTGCGCCAGCGGGGGCAACCCGCTGGGGTACTCTTACATGCT 36070

Qy        641 GTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTC 700
              ||||| |||||||||||||||||  || |||||||||||| ||||||| |||| |||
Db      36069 GTGGAGGCTTTGTTCTTTTGCTCTTTGCAGTAAATCTTGCTGTTGCTCATTTTTTGGGTC 36010

Qy        701 CACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGT 760
                |||||||||||||||||||||||||  | || ||||||  |||||| |||| ||||
Db      36009 TGTACTACCTTTATGAGCTGTAACACTCAGCGTGAAGGTCTGCAGCTTCACTCCTCAAGT 35950

Qy        761 CAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGC 820
              || ||   |    | || |  | ||||||||| ||||||| | ||||  ||||||||||
Db      35949 CAGCAATACCACGAACCCCCCGGAGGAACAAACAACTCCCAACGTGCCACCTTTAAGAGC 35890

Qy        821 TGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCC 880
              |||||||||||||||||||  |||||  ||| |||||||||||||||||||||||| |||||
Db      35889 TGTAACACTCACTGCGAAGGTCTGCGACTTTACTCCTGAAGTCAGTGAGACCACGAACCC 35830

Qy        881 ACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACAC 940
              |||||||||||||||||  |  ||   |||| ||||||||||||||| ||||||||||
Db      35829 ACCAGAAGGAAGAAACTCCAGGTACCTGTGAACATCTGAAGGAACAAACGCCAGACACAC 35770

Qy        941 CATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCA 1000
              || |||| ||| ||| ||||||||| || ||||||    ||||||||||||||||||||
Db      35769 CAGCTTTAAGAACTGCAACACTCACTGCGAGGGTC-CCAGCTTCATTCTTGAAGTCAGCG 35711

Qy       1001 AGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGA 1060
              |||||| ||||||||| |||||||| |||||||||||||||  |  || |  |  | |
Db      35710 AGACCAGGAACCCACCAGAAGGAACCAATTCCAGACACAGTACAATCTCAGCTTACTGCA 35651

Qy       1061 TCTGTGGCTTCCAGGGTT 1078
                |    || ||| |||||
Db      35650 ACCTCTGCCTCCTGGGTT 35633


RESULT 45
ABZ67757/c
ID   ABZ67757 standard; DNA; 41206 BP.
XX
AC   ABZ67757;
XX
DT   26-MAR-2003  (first entry)
XX
DE   Human secreted protein encoding genomic DNA SEQ ID NO 1280.
XX
KW   Human; secreted protein; nootropic; neuroprotective; cytostatic;
KW   virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW   vulnerary; antibacterial; antiparkinsonian; antisickling; antianaemic;
KW   antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW   antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;
KW   antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW   cardiovascular disorder; neurological disease; nephrotropic;
KW   gene therapy; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277186-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-US009188.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-040583/03.
XX
PT   New human secreted proteins encoded by genes contained in cDNA clones
PT   (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT   multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT   West Nile fever.
XX
PS   Disclosure; Page 1953-1963; 2423pp; English.
XX
CC   The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC   encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC   treating or ameliorating medical conditions e.g. by protein or gene
```

```
CC   therapy. The genes are isolated from a range of human tissues disclosed
CC   in the specification. The nucleic acids, proteins, antibodies and
CC   (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC   (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC   adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC   lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC   myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC   cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC   bacterial, fungal and parasitic infections
XX
SQ   Sequence 41206 BP; 11685 A; 9614 C; 9725 G; 10182 T; 0 U; 0 Other;

  Query Match            20.0%;  Score 400.8;  DB 10;  Length 41206;
  Best Local Similarity   80.7%;  Pred. No. 6.9e-86;
  Matches  499;  Conservative    0;  Mismatches  102;  Indels   17;  Gaps    2;

Qy        461 TAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAAT 520
              |||||||||| ||||||| ||||||||||||||||||||                ||||
Db      36233 TAGCTAAAGGTTTGTAAAGGCACCAATCAGCACTCTGTAA----------------AAAT 36190

Qy        521 GCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAG 580
              | |||||||||| |||||||||||||||||||||||||||||||| |||| ||||||||
Db      36189 GGACCAATCAGCGCTCTGTAAAATGGACCAATCAGCAGGATGTGGGTGGGGCCAAATAAG 36130

Qy        581 GGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACT 640
              ||| ||||| ||||||||| | |||||| | |||||||| || |||| | ||| ||  ||
Db      36129 GGAATAAAAGCTGGCCACCTGCGCCAGCGGGGGCAACCCGCTGGGGTACTCTTACATGCT 36070

Qy        641 GTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTC 700
              ||||| |||||||||||||||||  || |||||||||||| |||||||| |||| |||
Db      36069 GTGGAGGCTTTGTTCTTTTGCTCTTTGCAGTAAATCTTGCTGTTGCTCATTTTTTGGGTC 36010

Qy        701 CACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGT 760
                 |||||||||||||||||||||||||  | || ||||||  ||||||| |||| ||||
Db      36009 TGTACTACCTTTATGAGCTGTAACACTCAGCGTGAAGGTCTGCAGCTTCACTCCTCAAGT 35950

Qy        761 CAACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGC 820
              || ||   |    | || |  | |||||||||| ||||||| | ||||  ||||||||||
Db      35949 CAGCAATACCACGAACCCCCCGGAGGAACAAACAACTCCCAACGTGCCACCTTTAAGAGC 35890

Qy        821 TGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCC 880
              |||||||||||||||||| |||||  ||| |||||||||||||||||||||| |||||
Db      35889 TGTAACACTCACTGCGAAGGTCTGCGACTTTACTCCTGAAGTCAGTGAGACCACGAACCC 35830

Qy        881 ACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACAC 940
              |||||||||||||||||  |  ||   |||| ||||||||||||||| ||||||||||
Db      35829 ACCAGAAGGAAGAAACTCCAGGTACCTGTGAACATCTGAAGGAACAAACGCCAGACACAC 35770

Qy        941 CATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCA 1000
              || |||| ||| ||| |||||||||| || ||||||    ||||||||||||||||||
Db      35769 CAGCTTTAAGAACTGCAACACTCACTGCGAGGGTC-CCAGCTTCATTCTTGAAGTCAGCG 35711

Qy       1001 AGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGA 1060
              |||||| ||||||||| |||||||| ||||||||||||||||  |  || |  |  |  |
Db      35710 AGACCAGGAACCCACCAGAAGGAACCAATTCCAGACACAGTACAATCTCAGCTTACTGCA 35651

Qy       1061 TCTGTGGCTTCCAGGGTT 1078
                |    || ||| |||||
Db      35650 ACCTCTGCCTCCTGGGTT 35633


Search completed: January 22, 2006, 04:09:37
Job time : 830.081 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:39:03 ; Search time 245.225 Seconds
(without alignments)
14504.671 Million cell updates/sec

Title: US-09-728-552A-3_COPY_59000_61000
Perfect score: 2001
Sequence: 1 aaaggcttctgaaatcagac.....cttgaagggttagagggact 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	80595	3	US-09-078-294-3	Sequence 3, Appli
	2	1999.4	99.9	80246	3	US-09-078-294-4	Sequence 4, Appli
	3	490.6	24.5	25589	3	US-09-949-002-716	Sequence 716, App
c	4	457.4	22.9	44870	3	US-09-949-016-12554	Sequence 12554, A
c	5	457.4	22.9	44881	3	US-09-949-016-16349	Sequence 16349, A
	6	452.6	22.6	1859	3	US-08-691-563C-46	Sequence 46, Appl
	7	452.6	22.6	1859	3	US-09-374-766-46	Sequence 46, Appl
	8	452.6	22.6	1859	3	US-08-979-847B-42	Sequence 42, Appl
	9	436.6	21.8	144322	3	US-09-949-016-15316	Sequence 15316, A
c	10	435.8	21.8	89625	3	US-09-949-016-17012	Sequence 17012, A
c	11	432.8	21.6	360470	3	US-09-949-016-13173	Sequence 13173, A
c	12	421.6	21.1	84132	3	US-09-949-016-16241	Sequence 16241, A
	13	412.2	20.6	2834	3	US-09-305-384-6	Sequence 6, Appli
	14	412.2	20.6	6235	3	US-09-305-384-5	Sequence 5, Appli
	15	412.2	20.6	6235	3	US-09-525-160B-6	Sequence 6, Appli
	16	412.2	20.6	6679	3	US-09-305-384-1	Sequence 1, Appli
	17	412.2	20.6	6679	3	US-09-525-160B-5	Sequence 5, Appli
	18	411.8	20.6	115963	3	US-09-949-016-12298	Sequence 12298, A
c	19	407.2	20.3	190078	3	US-09-949-016-12707	Sequence 12707, A
c	20	407.2	20.3	190078	3	US-09-949-016-17026	Sequence 17026, A
	21	406.8	20.3	601	3	US-09-949-016-184449	Sequence 184449,
c	22	404.8	20.2	93364	3	US-09-949-016-14890	Sequence 14890, A
	23	403.8	20.2	190078	3	US-09-949-016-12707	Sequence 12707, A
	24	403.8	20.2	190078	3	US-09-949-016-17026	Sequence 17026, A
	25	401.2	20.0	2942	3	US-10-104-047-463	Sequence 463, App
c	26	392.2	19.6	107421	3	US-09-949-016-15532	Sequence 15532, A
c	27	382	19.1	80355	3	US-09-949-016-12735	Sequence 12735, A
c	28	382	19.1	80357	3	US-09-949-016-13572	Sequence 13572, A
c	29	379.4	19.0	601	3	US-09-949-016-184415	Sequence 184415,
c	30	372	18.6	421491	3	US-09-949-016-12805	Sequence 12805, A
c	31	372	18.6	421494	3	US-09-949-016-14060	Sequence 14060, A
	32	371	18.5	80706	3	US-09-949-016-15347	Sequence 15347, A
c	33	370.4	18.5	52992	3	US-09-949-016-16105	Sequence 16105, A
c	34	367.4	18.4	33578	3	US-09-949-016-15670	Sequence 15670, A
c	35	365.4	18.3	87644	3	US-09-949-016-16041	Sequence 16041, A
	36	361.8	18.1	1043	3	US-09-422-576D-6	Sequence 6, Appli
c	37	359.8	18.0	124700	3	US-09-949-016-11817	Sequence 11817, A
c	38	359.8	18.0	124701	3	US-09-949-016-15439	Sequence 15439, A
c	39	357.2	17.9	181429	3	US-09-949-016-12372	Sequence 12372, A
c	40	357.2	17.9	181430	3	US-09-949-016-15772	Sequence 15772, A
	41	355	17.7	94019	3	US-09-949-016-13203	Sequence 13203, A
c	42	350.2	17.5	88557	3	US-09-949-016-17028	Sequence 17028, A
	43	349.2	17.5	157644	3	US-09-949-016-16179	Sequence 16179, A
	44	349.2	17.5	157644	3	US-09-949-016-16180	Sequence 16180, A
c	45	348.6	17.4	22888	3	US-09-949-016-14659	Sequence 14659, A
	46	343.6	17.2	68444	3	US-09-949-016-13968	Sequence 13968, A
	47	342.6	17.1	235064	3	US-09-949-016-15390	Sequence 15390, A
	48	342	17.1	1091	3	US-09-422-576D-5	Sequence 5, Appli
	49	341.8	17.1	104077	3	US-09-949-016-13593	Sequence 13593, A
c	50	341.8	17.1	246230	3	US-09-949-016-17019	Sequence 17019, A
c	51	341.8	17.1	246230	3	US-09-949-016-17020	Sequence 17020, A
c	52	341.8	17.1	246230	3	US-09-949-016-17021	Sequence 17021, A
c	53	341.8	17.1	246230	3	US-09-949-016-17022	Sequence 17022, A
	54	341.2	17.1	65485	3	US-09-949-016-13757	Sequence 13757, A
	55	341.2	17.1	65552	3	US-09-949-016-12303	Sequence 12303, A
	56	340.2	17.0	54945	3	US-09-967-669-10	Sequence 10, Appl
c	57	338.8	16.9	100550	3	US-09-949-016-11835	Sequence 11835, A
c	58	338.8	16.9	100551	3	US-09-949-016-16207	Sequence 16207, A
c	59	337	16.8	69909	3	US-09-949-016-13423	Sequence 13423, A
c	60	335.6	16.8	135010	3	US-09-949-016-17234	Sequence 17234, A
c	61	334.6	16.7	36221	3	US-09-954-556-29	Sequence 29, Appl
c	62	334.4	16.7	150780	3	US-09-949-016-14711	Sequence 14711, A
	63	332.2	16.6	59140	3	US-09-949-016-16023	Sequence 16023, A
	64	332	16.6	75394	3	US-09-949-016-11754	Sequence 11754, A
c	65	329.6	16.5	91538	3	US-09-949-016-15703	Sequence 15703, A
	66	326.6	16.3	186959	3	US-09-949-016-13125	Sequence 13125, A
c	67	326.4	16.3	133278	3	US-09-949-016-12524	Sequence 12524, A
c	68	326	16.3	312474	3	US-09-949-016-17434	Sequence 17434, A
	69	323.8	16.2	24671	3	US-09-949-016-12842	Sequence 12842, A
	70	323.8	16.2	26274	3	US-09-949-016-12841	Sequence 12841, A
	71	323.8	16.2	26276	3	US-09-949-016-15026	Sequence 15026, A
	72	323.8	16.2	26276	3	US-09-949-016-15027	Sequence 15027, A
	73	323.8	16.2	26276	3	US-09-949-016-15028	Sequence 15028, A
	74	323.8	16.2	26276	3	US-09-949-016-15029	Sequence 15029, A
	75	322.2	16.1	96048	3	US-09-949-016-15966	Sequence 15966, A
	76	320.2	16.0	8399	3	US-09-573-080A-26	Sequence 26, Appl
	77	318.8	15.9	601	3	US-09-949-016-126727	Sequence 126727,
c	78	318.6	15.9	601	3	US-09-949-016-104912	Sequence 104912,
c	79	318.6	15.9	20441	3	US-09-949-016-14476	Sequence 14476, A
	80	318.4	15.9	938	3	US-09-719-554-18	Sequence 18, Appl
	81	317.8	15.9	601	3	US-09-949-016-126728	Sequence 126728,
c	82	317.6	15.9	601	3	US-09-949-016-192727	Sequence 192727,
	83	312.8	15.6	1722	3	US-08-691-563C-58	Sequence 58, Appl
	84	312.8	15.6	1722	3	US-09-374-766-58	Sequence 58, Appl
	85	312.8	15.6	1722	3	US-08-979-847B-54	Sequence 54, Appl
	86	312.4	15.6	1165	3	US-09-422-576D-25	Sequence 25, Appl
	87	312.4	15.6	1831	3	US-09-422-576D-1	Sequence 1, Appli
c	88	310.2	15.5	33155	3	US-09-949-016-16421	Sequence 16421, A
	89	308	15.4	121068	3	US-09-949-016-14138	Sequence 14138, A
c	90	306.6	15.3	239527	3	US-09-949-016-15980	Sequence 15980, A
	91	305	15.2	237241	3	US-09-949-016-16101	Sequence 16101, A
	92	303.4	15.2	767677	3	US-09-949-016-12147	Sequence 12147, A
	93	303.4	15.2	767677	3	US-09-949-016-17361	Sequence 17361, A
	94	302.4	15.1	945	3	US-09-719-554-14	Sequence 14, Appl
	95	301	15.0	801	3	US-09-422-576D-7	Sequence 7, Appli
	96	300.8	15.0	572	3	US-09-621-976-10590	Sequence 10590, A
	97	300.6	15.0	39498	3	US-09-949-016-12410	Sequence 12410, A

98	300.6	15.0	39498	3	US-09-949-016-16505	Sequence 16505, A
c 99	299.8	15.0	128779	3	US-09-497-855A-38	Sequence 38, Appl
c 100	299.4	15.0	601	3	US-09-949-016-155802	Sequence 155802,

ALIGNMENTS

RESULT 1
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match 100.0%; Score 2001; DB 3; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59000 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 59059

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59060 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 59119

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59120 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 59179

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59180 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 59239

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59240 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 59299

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59300 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 59359

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59360 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 59419

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59420 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 59479

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59480 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 59539

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59540 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 59599

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59600 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 59659

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59660 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 59719

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59720 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 59779

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59780 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 59839

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59840 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 59899

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59900 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 59959

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59960 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 60019

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60020 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 60079

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60080 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 60139

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60140 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 60199

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60200 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 60259

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60260 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 60319

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60320 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 60379

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60380 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 60439

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60440 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 60499

Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60500 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 60559

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60560 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 60619

Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60620 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 60679

Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      60680 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 60739

Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60740 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 60799

Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60800 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 60859

Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60860 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 60919

Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60920 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 60979

Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001
              |||||||||||||||||||||
Db      60980 CCTTGAAGGGTTAGAGGGACT 61000


RESULT 2
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

  Query Match             99.9%;  Score 1999.4;  DB 3;  Length 80246;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2000;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58767 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 58826

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58827 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 58886

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58887 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 58946

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58947 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 59006

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59007 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 59066

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59067 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCA 59126

Qy        361 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      59127 AATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTG 59186

Qy        421 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59187 GGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATG 59246

Qy        481 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59247 CACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTA 59306

Qy        541 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59307 AAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCC 59366

Qy        601 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59367 GAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 59426

Qy        661 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59427 CTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTG 59486

Qy        721 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59487 TAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACAGACCACGAACCCACT 59546

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59547 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 59606

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59607 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 59666

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59667 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 59726

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59727 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 59786

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59787 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 59846

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59847 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 59906

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59907 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 59966

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59967 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 60026

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60027 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 60086

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      60087 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACAAGGTC 60146

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60147 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 60206

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60207 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAGTACTCAGGAGGCTGAGGCAGGGG 60266
```

```
Qy       1501 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60267 AATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCA 60326

Qy       1561 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60327 TCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAAT 60386

Qy       1621 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60387 ATAATCAAGAATATTGACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACA 60446

Qy       1681 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60447 CTAAGTGTTTTACATGGATTAACTCATTTAATCTTAACAATAGCCCTATGAAGTCAGTGC 60506

Qy       1741 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60507 TGTTATTATCTCCACTTTATAGATAAGGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 60566

Qy       1801 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60567 TGGCCATGCTTGCATTCTCAGTTTTTGAAGCAACTGTTACAGGAATCTGGTGTGAGAAAT 60626

Qy       1861 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60627 GCTCTAACAAGATGTGAGTCAGGGGTTGGGAGGTACTGAGTCTGAGTTGGGCAGTTGGGG 60686

Qy       1921 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60687 ATGGAAGGATGGATGAAGAACAGCTTGACAGAGAAGCTGACACTTGGCAACTCTGTGGGA 60746

Qy       1981 CCTTGAAGGGTTAGAGGGACT 2001
              |||||||||||||||||||||
Db      60747 CCTTGAAGGGTTAGAGGGACT 60767
```

```
RESULT 3
US-09-949-002-716
; Sequence 716, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 716
;   LENGTH: 25589
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(25589)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-002-716

  Query Match             24.5%;  Score 490.6;  DB 3;  Length 25589;
  Best Local Similarity   66.4%;  Pred. No. 4.8e-119;
  Matches  868;  Conservative    0;  Mismatches  379;  Indels   60;  Gaps    9;

Qy        323 CCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCGCCAAATCCCAACAGCAGCTGGGGTG 382
              | | |  | |||    ||| |||   |     | |  ||  || | ||||| |  | ||
Db      14437 CACCAATCGGCACTCTGTATCTAGCTCAAGGTTTGTAAACACCAATCAGCACCCTGTGTC 14496

Qy        383 TCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGAC 442
```

```
Db      14497 TAGCTCAGGGTTTGTGACTGCACCAATCGACACTCTGCATCTAGCTACTCTGGTGGGGAC 14556

Qy        443 TTGGAGAACTTTTGTG-------TCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTC 495
              ||||||||| ||||||       | || |||    ||      | ||         ||
Db      14557 TTGGAGAACCTTTGTGTGGACACTGTATCTAGCTAATCTGGTGGGGACGTGGAGAACCTT 14616

Qy        496 TGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAA------------ 543
              |||||||||||| | |||| ||||| ||||||||||| | |||| |||
Db      14617 TGTGTCTAGCTGAGGGATGGTAAACGCACCAATCAGTGCCCTGTCAAAACAGACCACTGG 14676

Qy        544 --TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCG 601
                | |||||||||||||||||||||| |||| || ||||| || ||||| | |||  ||||
Db      14677 GCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCG 14736

Qy        602 AGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGC 661
              |||||||||||||||||  ||||||||| |||||||||||||||||||||||| ||| ||
Db      14737 AGCCAGCAGTGGCAACCGTCTCGGGTCCGCTTCCACACTGTGGAAGCTTTGTTGTTTCGC 14796

Qy        662 TCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGT 721
              ||||  |||||||||| |||||||||||| |||||| ||||||||| ||||||||||||||
Db      14797 TCTTTGCAATAAATCCTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGT 14856

Qy        722 AACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACT 780
              |||||||  |||| ||||||  ||||| |||||||| || | ||||||||| |||||
Db      14857 GACACTCATGGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACCACGAGCCCACC 14916

Qy        781 GGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGC 840
              | |||||| || |||||| || |  | |||||| ||||||||||||||||||||||||
Db      14917 AGGAGGAACGAACAACTCCAGACGCACCGCCTTT-AGAGCTGTAACACTCACTGCGAAGG 14975

Qy        841 TCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTG 900
              |||||||||||||||||| || ||| |||| ||| |||||| ||||| ||||||||||
Db      14976 TCTGCAGCTTCACTCCTG-AGCCAGCGAGAGCACCAACCCAACAGAAAGAAGAAACTCCA 15034

Qy        901 GACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACA 960
              ||||| | ||| ||| ||||||| |||||||||| | ||| | || ||||||||||||
Db      15035 AACACATCCGAACATCAGAAGGAAAAAACTCCAGACGCGCCA-CCTTAAGAGCTGTAACA 15093

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAA 1020
              |||||||| |||||||   |||||||||||||||||||   | ||||||||||||||
Db      15094 CTCACCGCGAGGGTCCACGGCTTCATTCTTGAAGTCAGTGACACCAAGAACCCACCAATT 15153

Qy       1021 GGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTAC 1080
              |      |   |  || ||  | |  | ||  |        |      | ||  |  ||
Db      15154 GCGGACACACCACCACCCATAATGCCACCTTGAAACCACTTTGTCTCCGTCAGGCCCTAA 15213

Qy       1081 TCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCAC 1140
              |  |   |||      |   |  || |||||      |     |  ||
Db      15214 GCTTCTTGGACAAGCGGAGAGAAGTTCAATGTGGGAACAGCAGGGAGGGCCAAACAG--- 15270

Qy       1141 ATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA 1200
                                        ||  | | ||  | ||     | | | |     |
Db      15271 -------------------------TTTTGTTTCATTTTGGTTTACTTTTAACCCTTATT 15305

Qy       1201 GCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGT 1260
              |       |   | || || || ||     | |||| | |  ||       |  || |||
Db      15306 ACAATCCTTATAACTAGTAAGTAGAAAAAGAATAAGAAAATGGTT------AACAGCAGT 15359

Qy       1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGC 1320
              ||  | | | |   ||              |  ||  | |||||  |    |||  |
Db      15360 GGTTTGACATAATAGGAAAAGCACTGTGCTTTAAAAAAAAAAAAAATTTCGGCCGGCCAT 15419

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTC 1380
              | |||||||  |||||||||||||| |||||||||||| ||| | |  ||||| ||||||
Db      15420 GGTGGCTCACACCTGTAATCCCAGCAATTTGGGAGGCCGAGGTGGGCAGATCATGAGGTC 15479

Qy       1381 AGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAA 1440
              ||| |||||||||||||||  | |||||||  |||||| || | ||||||||||||||||
Db      15480 AGGAGTTCGAGACCAGCCTGACTAACATGGCAAAACCCTGTTTTTACTAAAAATACAAAA 15539

Qy       1441 ATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGGCAGGG 1499
              |||||||  ||  || ||||||  | ||||||||  ||| |||||||  ||||||||||||
```

```
Db      15540 ATTAGCCAGGCATGCTGGCACATACCTGTAATCTCAGCTACTCAGTAGGCTGAGGCAGGA 15599

Qy       1500 GAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCC 1559
              ||||| |||||||   |||||||||||||||||||||||||| |||| |||| |||||||
Db      15600 GAATCACTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCAAGGTCACGCCACTGCACTCC 15659

Qy       1560 ATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAA 1606
              | |||||||  |||||| |||||||||||||||||||| ||| ||| |
Db      15660 AGCCTGGGCAACAGAGTGAGACTCTGTCTCAAAAAAATAAATAAATA 15706


RESULT 4
US-09-949-016-12554/c
; Sequence 12554, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12554
;   LENGTH: 44870
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12554

  Query Match            22.9%;  Score 457.4;  DB 3;  Length 44870;
  Best Local Similarity  67.4%;  Pred. No. 3.6e-110;
  Matches  781;  Conservative    0;  Mismatches  354;  Indels   23;  Gaps    9;

Qy        493 CTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAA--------- 543
              || |||||| |||| | |||||||||| ||||||||||||| |||| |||
Db       8432 CTTTGTGTCCAGCTCAGGGATTGTAAACGCACCAATCAGCACCCTGTCAAAACAGACCAC 8373

Qy        544 -----TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCAC 598
                  | |||||||||||||||||||| |||| || ||||| || ||||| |||||  |
Db       8372 TCGGTTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCTGGCTGC 8313

Qy        599 CCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTT 658
              || |||||||||||||||||  ||  |||||||||||||||  ||||||||||||||||||
Db       8312 CCCAGCCAGCAGTGGCAACCCGCTTAGGTCCCCTTCCACACCATGGAAGCTTTGTTCTTT 8253

Qy        659 TGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGC 718
               ||||||  ||||| |||||||| ||||||| |||||| |||||| || | |||||||||
Db       8252 CGCTCTTTGCAATACATCTTGCTACTGCTCACTCTTTGGGTCCACGCTGCTTTTATGAGC 8193

Qy        719 TGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCC 777
              |||||||||||| || | ||||||  |||||| |||| ||| || | ||||| ||| |||
Db       8192 TGTAACACTCACCGCAAAGGTCTGCAGCTTCACTCCTAAAGCCAGCAAGACCGCGAGCCC 8133

Qy        778 ACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGA 837
               | || |||||| || |||||| || | |||||| ||||||||||||||||||||| || |
Db       8132 CCAGGGAGGAACGAACAACTCCAGACGCGCTGCC-TTAAGAGCTGTAACACTCACCGCAA 8074

Qy        838 AGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACT 897
              || |||||||||||||| ||| || ||| ||||||| ||||||||||||||||||||||
Db       8073 AGGTCTGCAGCTTCACTTCTG-AGCCAGCGAGACCATGAACCCACCAGAAGGAAGAAACT 8015

Qy        898 CTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTA 957
              | | ||||| |  || ||| ||||||| |||||||||||| ||| | || |||||||||
Db       8014 CCGAACACATCCTAACATCAGAAGGAATAAACTCCAGACACGCCA-CCTTAAGAGCTGTA 7956

Qy        958 ACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCG 1017
              |||||||||| |||||||| |||||||||||||||||||  || |||||| ||||||
Db       7955 ACACTCACCGCGAGGGTCTGCGGCTTCATTCTTGAAGTCAGTGAGGCCAAGATCCCACCA 7896

Qy       1018 GAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGT 1077
                        |||||          | | | |   | | | |    | |
Db       7895 ATTCTGGACACACTTTTACACCACTATTTGTGTCCAATCAGGGTTTCTTCTCTACTTCCG 7836

Qy       1078 TACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAG 1137
              || | |  ||        ||  ||    | | | |       |   |       |
Db       7835 CACACAATGCAGCCTCTCTTCTCCTGACTACACAGAGCACACATTTCTTTATACTTCACG 7776

Qy       1138 CACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAAC 1197
              | |     | || |    |   |   || |     | ||||  | |         | ||
Db       7775 CTCTGTCCTGCATCGCATCGTTTCCTTCTAGCTTGACCCATACAGAAAAGAGCCAGAGAC 7716

Qy       1198 CTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGT 1257
                | || |          | | | |          |    || |  | |     |
Db       7715 TCAAATGACAAAGCCAAAATCAGAACTGTAGCGCCTATTTTAGCTCAGGAACTGGATGAC 7656

Qy       1258 AGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGG 1317
              ||   |  | ||   | | |  || |     | |       |      ||||||
Db       7655 AGAACTGTAAGCTAGAAAAAGCTGACTTCACCTTTACCATCATACCTTTAACAGGCCAGG 7596

Qy       1318 TGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCAC--G 1375
              || |||||||  |||||||||| |||||||||||||||||||| | |  ||||||  |
Db       7595 CACGGTGGCTCACACCTGTAATCCGAGCACTTTGGGAGGCCAAGGTGGGCAGATCACCTG 7536

Qy       1376 AGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATA 1435
              || |  || |||||||| ||||| || |||| |||| |||||||| ||||||||||||||
Db       7535 AGATTGGGAGTTCGAGCCCAGCGTGACCAATATGGAGAAACCCCATCTCTACTAAAAATA 7476

Qy       1436 CAAAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGG 1494
              | |||| ||||| || ||||||||||    |||||||||||| |||| ||||||||||
Db       7475 C-AAAAGTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGGAGGCTGAGG 7417

Qy       1495 CAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGC 1554
              |||| ||||| ||| |||  |||||||||||||||||||||| | |||   |||  ||
Db       7416 CAGGAGAATCGCTTTAACCCGGGAGGCAGAGGTTGCAGTGAGCCCACATCGTGCCATTGC 7357

Qy       1555 ACTCCATCCTGGGTGAC-AGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAGGAAA 1613
              |||||| ||||||  || |||| || |||| | ||||||| ||| |||  || ||  |||
Db       7356 ACTCCAGCCTGGGCAACAAGAGTGAAACTCCGCCTCAAAACAAACAAACGAACAAAAAAA 7297

Qy       1614 GGAAAATATAATCAAGAA 1631
                | || |   | |  ||
Db       7296 CCACAACACCTTAACCAA 7279


RESULT 5
US-09-949-016-16349/c
; Sequence 16349, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16349
;   LENGTH: 44881
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16349
```

```
  Query Match             22.9%;  Score 457.4;  DB 3;  Length 44881;
  Best Local Similarity   67.4%;  Pred. No. 3.6e-110;
  Matches  781;  Conservative    0;  Mismatches  354;  Indels   23;  Gaps    9;

Qy        493 CTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAA--------- 543
              || |||||| |||| | |||||||||| |||||||||||||| |||| |||
Db       8439 CTTTGTGTCCAGCTCAGGGATTGTAAACGCACCAATCAGCACCCTGTCAAAACAGACCAC 8380

Qy        544 -----TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCAC 598
                  | |||||||||||||||||||| |||| || ||||| || ||||| |||||   |
Db       8379 TCGGTTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCTGGCTGC 8320

Qy        599 CCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTT 658
              || |||||||||||||||||| || ||||||||||||||   ||||||||||||||||||
Db       8319 CCCAGCCAGCAGTGGCAACCCGCTTAGGTCCCCTTCCACACCATGGAAGCTTTGTTCTTT 8260

Qy        659 TGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGC 718
              ||||||  ||||| |||||||| ||||||| |||||| |||||| || | |||||||||
Db       8259 CGCTCTTTGCAATACATCTTGCTACTGCTCACTCTTTGGGTCCACGCTGCTTTTATGAGC 8200

Qy        719 TGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCC 777
              |||||||||||| || | ||||||  |||||| |||| ||| || | ||||| ||| |||
Db       8199 TGTAACACTCACCGCAAAGGTCTGCAGCTTCACTCCTAAAGCCAGCAAGACCGCGAGCCC 8140

Qy        778 ACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGA 837
              | || |||||| || |||||| || | |||||| ||||||||||||||||||| || |
Db       8139 CCAGGGAGGAACGAACAACTCCAGACGCGCTGCC-TTAAGAGCTGTAACACTCACCGCAA 8081

Qy        838 AGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACT 897
              || |||||||||||||| ||| || ||| ||||||| |||||||||||||||||||||||
Db       8080 AGGTCTGCAGCTTCACTTCTG-AGCCAGCGAGACCATGAACCCACCAGAAGGAAGAAACT 8022

Qy        898 CTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTA 957
              | | ||||| |  || ||| ||||||| |||||||||||| ||| | || |||||||||
Db       8021 CCGAACACATCCTAACATCAGAAGGAATAAACTCCAGACACGCCA-CCTTAAGAGCTGTA 7963

Qy        958 ACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCG 1017
              |||||||||| |||||||| ||||||||||||||||||||  || |||||| ||||||
Db       7962 ACACTCACCGCGAGGGTCTGCGGCTTCATTCTTGAAGTCAGTGAGGCCAAGATCCCACCA 7903

Qy       1018 GAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGT 1077
                  |      ||||         | |  | |   | | | |    | |      |
Db       7902 ATTCTGGACACACTTTTACACCACTATTTGTGTCCAATCAGGGTTTCTTCTCTACTTCCG 7843

Qy       1078 TACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAG 1137
              || | |  ||       ||  ||   |  | | |        |      |       |
Db       7842 CACACAATGCAGCCTCTCTTCTCCTGACTACACAGAGCACACATTTCTTTATACTTCACG 7783

Qy       1138 CACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAAC 1197
              | |      | || |    |   |   || |     | ||||  | |          | ||
Db       7782 CTCTGTCCTGCATCGCATCGTTTCCTTCTAGCTTGACCCATACAGAAAAGAGCCAGAGAC 7723

Qy       1198 CTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGT 1257
                |  ||  |       |  | |  |         |    || |   | |     |
Db       7722 TCAAATGACAAAGCCAAAATCAGAACTGTAGCGCCTATTTTAGCTCAGGAACTGGATGAC 7663

Qy       1258 AGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGG 1317
              ||   |   | ||   | | |  || |       | |         |     ||||||
Db       7662 AGAACTGTAAGCTAGAAAAAGCTGACTTCACCTTTACCATCATACCTTTAACAGGCCAGG 7603

Qy       1318 TGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCAC--G 1375
                || |||||||  |||||||||| |||||||||||||||||||| | |  ||||||  |
Db       7602 CACGGTGGCTCACACCTGTAATCCGAGCACTTTGGGAGGCCAAGGTGGGCAGATCACCTG 7543

Qy       1376 AGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATA 1435
              || |  || ||||||| ||||| || |||| |||| |||||||| ||||||||||||||||
Db       7542 AGATTGGGAGTTCGAGCCCAGCGTGACCAATATGGAGAAACCCCATCTCTACTAAAAATA 7483

Qy       1436 CAAAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGG 1494
              | |||| ||||| || ||||||||||    |||||||||||| ||||  ||||||||||
Db       7482 C-AAAAGTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGGAGGCTGAGG 7424

Qy       1495 CAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGC 1554
```

```
              |||| ||||| ||| |||  |||||||||||||||||||||||| | |||   |||  ||
Db       7423 CAGGAGAATCGCTTTAACCCGGGAGGCAGAGGTTGCAGTGAGCCCACATCGTGCCATTGC 7364

Qy       1555 ACTCCATCCTGGGTGAC-AGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAGGAAA 1613
              |||||| ||||||  || |||| || |||| | ||||||| ||| |||  || ||  |||
Db       7363 ACTCCAGCCTGGGCAACAAGAGTGAAACTCCGCCTCAAAACAAACAAACGAACAAAAAAA 7304

Qy       1614 GGAAAATATAATCAAGAA 1631
                | || |   | |  ||
Db       7303 CCACAACACCTTAACCAA 7286
```

RESULT 6
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
;   APPLICANT: Herve PERRON
;   APPLICANT: Frederic BESEME
;   APPLICANT: Frederic BEDIN
;   APPLICANT: Glaucia PARANHOS-BACCALA
;   APPLICANT: Florence KOMURIAN-PRADEL
;   APPLICANT: Colette JOLIVET
;   APPLICANT: Bernard MANDRAND
;   TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;   TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;   TITLE OF INVENTION: THERAPEUTIC PURPOSES
;   NUMBER OF SEQUENCES: 92
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Oliff & Berridge
;   STREET: 700 South Washington Street, Suite 300
;   CITY: Alexandria
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 22314
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/691,563C
;   FILING DATE: 02-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Berridge, William P.
;   REGISTRATION NUMBER: 30,024
;   REFERENCE/DOCKET NUMBER: WPB 38588
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 703-836-6400
;   TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 46:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1859 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-08-691-563C-46

```
  Query Match             22.6%;  Score 452.6;  DB 3;  Length 1859;
  Best Local Similarity   91.6%;  Pred. No. 1.6e-109;
  Matches  490;  Conservative    0;  Mismatches   44;  Indels    1;  Gaps    1;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||||| |||||||||| || ||||||||||||||||||||||
Db       1325 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAATTCCTATACCAACCTCTGGAGTTGGGC 1384

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||| |||| ||||||||| ||||  || | |||||||||||| || |||| |||||
Db       1385 AACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGG 1444

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||| ||||||||||||||||||||||||
```

```
Db       1445 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCT 1504

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||| |||||||| |||||||||
Db       1505 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACC 1564

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||| |||| |||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db       1565 CCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 1624

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              ||||||||||||| |||| |||||||||||| |||||||||||| ||| ||||| | ||
Db       1625 CACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGTCATTGCCC 1684

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||| |||||||||||||||| |||| ||| ||||||||||||||||||||||||
Db       1685 AATTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGTGAAGCCAGCT 1744

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              || |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1745 GGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 1804

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCAC 534
              ||| ||||||| |||||||||||||||||||||||||||||| ||||||||||||
Db       1805 GCAACAATCAGTGCTCTGTGTCTAGCTAAAGGATTGTAAATACACCAATCAGCAC 1859


RESULT 7
US-09-374-766-46
; Sequence 46, Application US/09374766
; Patent No. 6579526
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/374,766
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-374-766-46

  Query Match             22.6%;  Score 452.6;  DB 3;  Length 1859;
  Best Local Similarity   91.6%;  Pred. No. 1.6e-109;
  Matches  490;  Conservative    0;  Mismatches   44;  Indels    1;  Gaps    1;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              |||||||||||||||||||||| |||||||||| || ||||||||||||||||||||||
Db       1325 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAATTCCTATACCAACCTCTGGAGTTGGGC 1384

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||| |||| |||||||||| ||||  || | |||||||||||| || |||| |||||
Db       1385 AACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGG 1444

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||| ||||||||||||||||||||||||
Db       1445 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCT 1504

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||| |||||||| |||||||||
Db       1505 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACC 1564

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||| |||| |||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db       1565 CCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 1624

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              ||||||||||||| |||| |||||||||||| |||||||||||| ||| ||||| | ||
Db       1625 CACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGTCATTGCCC 1684

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||| |||||||||||||||| |||| ||| ||||||||||||||||||||||||
Db       1685 AATTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGTGAAGCCAGCT 1744

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              || |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1745 GGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 1804

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCAC 534
              ||| ||||||| |||||||||||||||||||||||||||||| ||||||||||||
Db       1805 GCAACAATCAGTGCTCTGTGTCTAGCTAAAGGATTGTAAATACACCAATCAGCAC 1859


RESULT 8
US-08-979-847B-42
; Sequence 42, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYI
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1859 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-08-979-847B-42

  Query Match             22.6%;  Score 452.6;  DB 3;  Length 1859;
  Best Local Similarity   91.6%;  Pred. No. 1.6e-109;
  Matches  490;  Conservative    0;  Mismatches   44;  Indels    1;  Gaps    1;

Qy          1 AAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATACCAACCTCTGGAGTTGGGC 60
              ||||||||||||||||||||||| |||||||||| || ||||||||||||||||||||||
Db       1325 AAAGGCTTCTGAAATCAGACAACGCCTTTCAAATTCCTATACCAACCTCTGGAGTTGGGC 1384

Qy         61 GACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTAATAGTCGCATTTGG 120
              |||||| |||| |||||||||| ||||  || | |||||||||||| || |||| |||||
Db       1385 AACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGG 1444

Qy        121 GCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCT 180
              |||||||||||||||||| | |||||||||||||| ||||||||||||||||||||||||
Db       1445 GCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCT 1504

Qy        181 ACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACC 240
              ||||||| |||||||||| |||||||||||||||||||| |||||||| |||||||||
Db       1505 ACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACC 1564

Qy        241 CCTGGACCGACCCGCTGGCCCTTTCAATGGCCTAAAGAGCTCCCCTCTGGAGGACACTAC 300
              ||| |||| |||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db       1565 CCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTTCCCCTCTGGAGGACACTAC 1624

Qy        301 CACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGTCATCG-CC 359
              |||||||||||||| ||||  |||||||||| |||||||||||| ||| ||||| | ||
Db       1625 CACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGTCATTGCCC 1684

Qy        360 AAATCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCT 419
              || |||||| |||||||||||||||| |||| ||| ||||||||||||||||||||||||
Db       1685 AATTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGTGAAGCCAGCT 1744

Qy        420 GGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 479
              || |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1745 GGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAAT 1804

Qy        480 GCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCAC 534
              ||| |||||||  |||||||||||||||||||||||||||| ||||||||||||
Db       1805 GCAACAATCAGTGCTCTGTGTCTAGCTAAAGGATTGTAAATACACCAATCAGCAC 1859


RESULT 9
US-09-949-016-15316
; Sequence 15316, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15316
;   LENGTH: 144322
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(144322)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15316

  Query Match             21.8%;  Score 436.6;  DB 3;  Length 144322;
  Best Local Similarity   88.9%;  Pred. No. 1.9e-104;
  Matches  506;  Conservative    0;  Mismatches   59;  Indels    4;  Gaps    3;

Qy        476 AAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACT 535
              ||| |||||||||||||||||||||||||||||||||||| |||||| |||||| |||||||
Db      15433 AAACGCACCAATCAGCACTCTGTGTCTAGCTAAAGGACTGTAAACGCACCAGTCAGCACT 15492

Qy        536 CTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGC 595
              |||||||||||||| |||||||||||||||||| |||| ||||||||| | ||||| |||||
Db      15493 CTGTAAAATGGACTAATCAGCAGGATGTGGGTGGGGCCAAATAAGGCAATAAAAGCTGGC 15552

Qy        596 CACCCGAGCCAGCAG-TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTT 654
              || |  |||||||||  |||||||||||||||| ||||||||  |||||||||||||||
Db      15553 CAGCGAAGCCAGCAGCAGGCAACCCACTCGGGTGCCCTTCCATGCTGTGGAAGCTTTGTT 15612

Qy        655 CTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTAT 714
              |||| |||||||||| ||||||||||||||||||  || || |||| |||||||| ||
Db      15613 CTTTCGCTCTTCACAGTAAATCTTGCTGCTGCTCA--CTCTGGGTCCGCACTACCTTCAT 15670

Qy        715 GAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAACA-GACCACGA 773
              |||||||||||||| |||||||||||| |||| ||||||||||||||| || |||||| |
Db      15671 GAGCTGTAACACTCAATGCGAGGGTCTGCGGCTGCATTCCTGAAGTCAGCAAGACCACAA 15730

Qy        774 ACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACT 833
              |||||| || |||||||||| |||||| |||| ||  ||||||||||||||||||||||||
Db      15731 ACCCACCGGGAGGAACAAACAACTCCAGATGAGCCACCTTTAAGAGCTGTAACACTCACT 15790

Qy        834 GCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGA 893
              |||||| | ||| |||||||||||||||||||||||||||||| |||||||  |||||||||
Db      15791 GCGAAGGTGTGCGGCTTCACTCCTGAAGTCAGTGAGACCACGAACCCACTGGAAGGAAGA 15850

Qy        894 AACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGC 953
              ||||| ||||||| ||||| |||||||||||||||||| ||||||||||||||| |||||
Db      15851 AACTCCGGACACATCTGAACATCTGAAGGAACAAACTCCGGACACACCATCTTTAAGAGC 15910

Qy        954 TGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCC 1013
              |||||| |||||||  | |||| |||| |||||||||||||||||| |||||||| | ||
Db      15911 TGTAACGCTCACCGTGAAGGTCCGTGGTTTCATTCTTGAAGTCAGCGAGACCAAGCATCC 15970

Qy       1014 ACCGGAAGGAACAAATTCCAGACACAGTA 1042
              |||  |||||| |||||||||| |||||
Db      15971 ACCAAAAGGAATAAATTCCAGACGCAGTA 15999


RESULT 10
US-09-949-016-17012/c
; Sequence 17012, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17012
;   LENGTH: 89625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(89625)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17012

  Query Match            21.8%;  Score 435.8;  DB 3;  Length 89625;
  Best Local Similarity  66.4%;  Pred. No. 2.5e-104;
  Matches  872;  Conservative    0;  Mismatches  287;  Indels  154;  Gaps   11;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||||||  ||||||||||||||||||||||| | ||||||||||||||||||||||
Db      38298 TCCAGTGGGGAGGTGGAGAACTTTTGTGTCTAGCTCAGGGATTGTAAATGCACCAATCAG 38239

Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCA 550
              |||  |||                    |||   |||||||||| ||||||||||||||||||
Db      38238 CACCCTGACA----------------AAACAGACCAATCAGCTCTCTGTAAAATGGACCA 38195

Qy        551 ATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAG 610
              |||||||||||||||| |||| || ||||| || ||||| | |||  || ||||||||||
Db      38194 ATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCTGAGCCAGCAG 38135

Qy        611 TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAA 670
              ||||||||  || ||||||||||||| |||||||||||||||||| | ||||||  ||
Db      38134 TGGCAACCTGCTGGGGTCCCCTTCCATACTGTGGAAGCTTTGTTCTCTCGCTCTTTGCAG 38075

Qy        671 TAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCAC 730
              |||||||||||||||||| |||||| ||||||||| ||||||||||||||||||||||||
Db      38074 TAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTAACACTCAC 38015

Qy        731 TGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAAC 789
              || || ||||||  |||||| |||||||| || | ||||||||||||||| | |||||||
Db      38014 TGGGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACCACGAACCCACCGCAAGGAAC 37955

Qy        790 AAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCT 849
               || ||||||  ||  ||| ||| |||||||||| |||||||||  || ||| || ||||||
Db      37954 GAACAACTCCAGACATGCCGCC-TTAAGAGCTGCAACACTCACCGCAAAGGTCCGCAGCT 37896

Qy        850 TCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCT 909
              ||||||||| || ||  |||||||| |||| |||||||||||||||| | ||||| ||
Db      37895 TCACTCCTG-AGCCAACGAGACCACGAACCTACCAGAAGGAAGAAACTCCGAACACATCT 37837

Qy        910 GAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCA 969
              ||| ||| ||||||||||||||||||||  || |||| ||| |||||||||| |||||
Db      37836 GAACATCAGAAGGAACAAACTCCAGACACGACACCTTTAAGAACTGTAACACTTACCGCG 37777

Qy        970 AGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC-------GGAAGG 1022
              ||||||  |||||||||||||||||||  ||||||||||||||||         | | |
Db      37776 AGGGTCCATGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACCAATTCCGGATACG 37717

Qy       1023 AACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTT------------ 1070
                 ||     ||| | | | |||| | | |||||  | | |   ||
Db      37716 CTACCATGTTTTACAGACTCGAAAATATCTTTTTTTTTTTTTTTTTTACCTTTAACATG 37657

Qy       1071 -------------CCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAA 1117
                          |  | | | | || | ||  | ||  ||||  |   ||
Db      37656 TTTGAGATCTGGACTCAGCATTCCACAATTGTTATTGGCTAAGCAGCAGTTATGGCAAAG 37597

Qy       1118 CAGAGAATGGTTTGGAGGAGCACATGTGGGAATTGTTAT---------------------- 1156
              | | | ||          |   |  ||  | |  | ||
Db      37596 CTGTCATTGCCAAACGTCATAGCTGTTGACACTGATGATACTTCCAGTGGCTATGGGCAT 37537

Qy       1157 ---GGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCTGGAAGCAGGG 1213
                   |||     |||||     |||     || |||     |||| ||       |||| |
Db      37536 TGGTGACACATGTTGAGTGTGCCATGTTAAATGTCTTTAAAAACTTATGATTTAGCACTG 37477

Qy       1214 CCAGGAAATATAA--TCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTTGCATCTG 1271
                | ||||  | |   |  || ||  ||          ||   | | |    ||  || | |
Db      37476 AGAAGAAAAGTTACTGCACAGAAAAGCACAAAAGCAGGAGCATGGCATGTATTTGATGTT 37417

Qy       1272 AGACATGTAGATTATCA------------------------------------------- 1288
              ||  |   | ||  | |
Db      37416 AGTGAAACAAATATTTATAAGACTATTAAACAAAAGCAAGTCAATGTTTTCTTGCAAAGC 37357

Qy       1289 ------------------------------AGCAATTAATTAGAAAAAATATAGCCAGG 1317
                                           || ||  ||     ||||| |||     |
Db      37356 AACAACCAAGCCCTCTAATGAAACCCACAATAGGAAGAAACCTTTAAAAAGATAAAGCTG 37297

Qy       1318 TGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCAC--G 1375
                | |||||||  |||||||||||||||||||||||||||  || || |  ||||||  |
Db      37296 GCTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGTGGGGCAGATCACCTG 37237

Qy       1376 AGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATA 1435
              |||||||| |||| |||| |||||| |||||||||||||||||| ||||||| |||||||
Db      37236 AGGTCAGGAGTTCAAGACAAGCCTGACCAACATGGTGAAACCCCATCTCTACAAAAAATA 37177

Qy       1436 CAAAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGG 1494
              |||||| |||||  |  |||||||||    |||||||||||| ||||||||||||||||||
Db      37176 CAAAAAGTAGCCGAGCATGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGG 37117

Qy       1495 CAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGC 1554
              |||| ||||| ||||||| |||||| |||||||||||||||||| ||||  |||||| ||
Db      37116 CAGGAGAATCGCTTGAACCTGGGAGACAGAGGTTGCAGTGAGCCGAGATTGCACCACTGC 37057

Qy       1555 ACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAA 1607
              |||||| ||||||  ||||||| ||||||| ||||||| ||| ||||| |||||
Db      37056 ACTCCAGCCTGGGAAACAGAGCAAGACTCCGTCTCAATAAACAAAAACAAAAA 37004


RESULT 11
US-09-949-016-13173/c
; Sequence 13173, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13173
;   LENGTH: 360470
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13173

  Query Match            21.6%;  Score 432.8;  DB 3;  Length 360470;
  Best Local Similarity  84.8%;  Pred. No. 2.8e-103;
  Matches  519;  Conservative    0;  Mismatches   67;  Indels   26;  Gaps    2;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     233842 TCTAGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 233783
```

```
Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCA---------------- 533
              ||||||||||||||||| |||| ||||||| ||||||||||||
Db     233782 CACTCTGTGTCTAGCTCAAGGTTTGTAAACGCACCAATCAGCAACCTGTGAAAACGGACC 233723

Qy        534 --------CTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGT 585
                    |||||||||||| ||||| |||||||||||| | ||||| |||||||| |
Db     233722 AATCAGCTCTCTGTAAAATGGGCCAATTGGCAGGATGTGGGTGAGGTCAGATAAGGGAAT 233663

Qy        586 AAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGA 645
              |||| | ||| ||| ||||||||| |||||| || |||||||| |||| |||||||
Db     233662 AAAAGCAGGCTGCCCAAGCCAGCAGCAGCAACCTGCTGGGGTCCCCCTCCATGCTGTGGA 233603

Qy        646 AGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACAC 705
              |||||||||||||   |||||||||||||||||||||||||||||| |||||| ||||    |
Db     233602 AGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCGTGC 233543

Qy        706 TACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC- 764
              | |||||| |||||||||||||||| | || ||||||  |||||| |||||||| || |
Db     233542 TGCCTTTAAGAGCTGTAACACTCACGGTGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCA 233483

Qy        765 AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTA 824
              |||||||||||||||||||||||||  || ||||||  |  ||||  ||||||||||||||
Db     233482 AGACCACGAACCCACTGGAAGGAATGAACAACTCCAGCCGTGCCACCTTTAAGAGCTGTG 233423

Qy        825 ACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCCACCA 884
              |||||||||| |||| ||||||||||||||||||||||||||||| ||| |||||||||
Db     233422 ACACTCACTGTGAAGGTCTGCAGCTTCACTCCTGAAGTCAGTGAGATCACGAACCCACCA 233363

Qy        885 GAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATC 944
              |||||||||| |||||||||| ||||| |||||||||||||||||| ||||||||||
Db     233362 GAAGGAAGAAATTCTGGACACATCTGAACATCTGAAGGAACAAACTCCGGACACACCATC 233303

Qy        945 TTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGAC 1004
              ||| ||| |||||||||||||| || |||||| | |||||||||||||||||||| ||||
Db     233302 TTTAAGAACTGTAACACTCACTGCGAGGGTCCGCGGCTTCATTCTTGAAGTCAGCGAGAC 233243

Qy       1005 CAAGAACCCACC 1016
              |||||| ||||
Db     233242 CAAGAACTCACC 233231
```

```
RESULT 12
US-09-949-016-16241/c
; Sequence 16241, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16241
;   LENGTH: 84132
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(84132)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16241

  Query Match            21.1%;  Score 421.6;  DB 3;  Length 84132;
  Best Local Similarity  68.4%;  Pred. No. 1.4e-100;
```

```
  Matches  812;  Conservative    0;  Mismatches  214;  Indels  162;  Gaps   10;

Qy        437 GGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCT 496
              | | || |||||||||||||||||||| | |||||||||| ||||||||||||||| ||
Db     21963 GAGAACATGGAGAACTTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCACCCT 21904

Qy        497 GTGT--------CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGAC 548
              ||         | | | |     |   |||   |||||||||| ||||||||||||||
Db     21903 GTCAAAACGGACCAATCAACTCTCTGTAAAACAGACCAATCAGCTCTCTGTAAAATGGAC 21844

Qy        549 CAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGC 608
              |||||||||||||||| | |||| || ||||| || ||||||| |||  || ||||||||
Db     21843 CAATCAGCAGGATGTGCGTGGGGCCAGATAAGAGAATAAAAACAGGCTGCCTGAGCCAGC 21784

Qy        609 AGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCAC 668
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db     21783 AGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTTGC 21724

Qy        669 AATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTC 728
              |||||||||||| ||||||| |||||| ||||||||| ||||||||||||||||||||
Db     21723 AATAAATCTTGCTACTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTAACACTC 21664

Qy        729 ACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGA 787
              || || | |||||| |||||| |||||||| || | | |||||||||||||| || ||||
Db     21663 ACCGCAAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAAACCACGAACCCACCGGGAGGA 21604

Qy        788 ACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAG 847
              || || |||||| || | || ||| |||||||||||||||||||| |||||| |||||||
Db     21603 ACGAACAACTCCAGACGCGCCGCC-TTAAGAGCTGTAACACTCACCGCGAAGGTCTGCAG 21545

Qy        848 CTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACAC 907
              |||||||||| || ||| |||||||| |||||||||||||||||||| | |||||
Db     21544 CTTCACTCCTG-AGCCAGCGAGACCACGAACCCACCAGAAGGAAGAAACTCCGAACACAA 21486

Qy        908 CTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTG-TAACACTCACC 966
              | ||| ||| |||||||||||||| ||||| ||  |||| ||| ||| ||||||||||
Db     21485 CCGAACATCAGAAGGAACAAACTCCGGACACGCCGCCTTTAAGAACTGTTAACACTCACC 21426

Qy        967 GCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACA 1026
              |||||||| |||||||||||||||||||||  ||||||||||||||||
Db     21425 GCAAGGGTCCGTGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACC---------- 21376

Qy       1027 AATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACTCCAGT 1086
              |||||| ||||||  |  ||   | ||
Db     21375 AATTCCGGACACAAGAACAACAGTATAA------------------------------- 21348

Qy       1087 CATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACATGTGG 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21347 ------------------------------------------------------------ 21348

Qy       1147 GAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCTGGA 1206
              ||||||||||||||||||||||||||||||||||||||||||||||| |||    | |  | |
Db     21347 ----------------------------------AGGTTTTGACTTTATCCCCATAGG 21324

Qy       1207 AGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTTGC 1266
              | ||   |   || | |   | | | |       | || |  | | ||       | ||
Db     21323 CACTGGAAAACCTAACAGAGAATGATGTAACATTGTATTTTTAAAATTAAATCCAGTGGC 21264

Qy       1267 ATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCGATGGC 1326
              ||   |    |  | | |  |   ||      |||          ||  || | | ||||
Db     21263 ATGGCACCATTAGGGTTGACTAGAAAAGCTGGGAGA--------GGCTGGGCGTGGTGGC 21212

Qy       1327 TCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCAGGCGT 1386
              ||| |||||||||||||||||||||||||||  ||| | | ||||||| |||||||| ||
Db     21211 TCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACAAGGTCAGGAGT 21152

Qy       1387 TCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGC 1446
              || |||||||||||  |||| |||||||||||| |||| |||| |||| ||||||||||||
Db     21151 TCAAGACCAGCCTGATCAACGTGGTGAAACCCTGTCTGTACTTAAAACACAAAAATTAGC 21092

Qy       1447 CTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGGCAGGGGAATCT 1505
              | ||||||| ||| |||||||||||||||| ||||||||||||||||||    ||||
Db     21091 CGGGTGTGGCAGCATGCATCTGTAATCCCAGCTACTCAGGAGGCTGAGGCA---GAATTG 21035
```

```
Qy       1506 CTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCATCCTG 1565
             ||||||| |||||   ||||||||| |||||||||||     |||| |||||||| ||||
Db      21034 CTTGAACACGGGAGATGGAGGTTGCAATGAGCCAAGATTGTGCCACTGCACTCCAGCCTG 20975

Qy       1566 GGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAA 1613
             || || | ||||||| ||  ||| |||||||||||||| ||| |  | |
Db      20974 GGCGATATAGCGAGAGTCCATCTTAAAAAAAAAAAAAAGAAAGAAAAGA 20927


RESULT 13
US-09-305-384-6
; Sequence 6, Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
;  APPLICANT: Treco, Douglas A.
;  APPLICANT: Heartlein, Michael W.
;  APPLICANT: Selden, Richard F
;  TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
;  FILE REFERENCE: 07236/017001
;  CURRENT APPLICATION NUMBER: US/09/305,384
;  CURRENT FILING DATE: 1999-05-05
;  EARLIER APPLICATION NUMBER: US 60/084,649
;  EARLIER FILING DATE: 1998-05-07
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 2834
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-305-384-6

  Query Match             20.6%;  Score 412.2;  DB 3;  Length 2834;
  Best Local Similarity   65.9%;  Pred. No. 9.3e-99;
  Matches  805;  Conservative    0;  Mismatches  358;  Indels   58;  Gaps   12;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTC---------TAGCTAAAGGATTGTAAATGC 481
              || |||||||  |||||||||  ||||||||         || |||    ||       |
Db       1370 TCTGGTGGGGCCTTGGAGAATGTTTGTGTCGACACTCTGTATCTAGTTAATCTAGTGGGG 1429

Qy        482 ACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAA 541
              ||          || |||||||||| | ||||||||| |||||||||||| | |||| |
Db       1430 ACGTGGAGAACCTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCGCCCTGTCA 1489

Qy        542 AA--------------TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAA 587
              ||              | ||||||||||||||||||| |||| || ||||| || |||
Db       1490 AAACAGACCACTCGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAA 1549

Qy        588 AAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAG 647
              || | |||  |||||||||||||||||||| | | | ||||||  |||||| | ||| ||
Db       1550 AAGCAGGCTGCCCGAGCCAGCAGTGGCAACGCGCACAGGTCCCTATCCACAATATGGCAG 1609

Qy        648 CTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTA 707
              ||||||||||||||  ||  | |||||||||||| ||||||  | |||| |||||||||
Db       1610 CTTTGTTCTTTTGCTGTTTGCGATAAATCTTGCTACTGCTCGCTTTTTGGGTCCACACTG 1669

Qy        708 CCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AG 766
              | |||||||||||||||||||||  ||| ||||||  |||||| |||||||| ||   ||
Db       1670 CTTTTATGAGCTGTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCACTAAG 1729

Qy        767 ACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAAC 826
              ||||||| ||||| || |||||  || |||||| |  | |||||| |||||||| ||||
Db       1730 ACCACGAGCCCACCGGGAGGAATGAACAACTCCGGCCGCGCTGCC-TTAAGAGCTATAAC 1788

Qy        827 ACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGA 886
              |||||| |||||| ||||||||||||||||||  || ||| |||||||| ||||||||||
Db       1789 ACTCACCGCGAAGGTCTGCAGCTTCACTCCT-CAGCCAGCGAGACCACGAACCCACCAGA 1847

Qy        887 AGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTT 946
              ||||||||||  | ||||| ||||| ||| ||||||||||||||||||  ||||| | |
Db       1848 AGGAAGAAACTGCGAACACATCTGAACATCAGAAGGAACAAACTCCAGATGCACCA-CCT 1906

Qy        947 TCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCA 1006
              | ||||||||||||||||| || |||||| | |||||| ||||||||||||  ||||||
Db       1907 TAAGAGCTGTAACACTCACTGCGAGGGTCCGCGGCTTCCTTCTTGAAGTCAGTGAGACCA 1966

Qy       1007 AGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTG 1066
              || || |||| |          |   | |  | | |                      ||
Db       1967 AGCACTCACCAGTTTCGGACACAAGCCCAGGAGTTTGAGATCAGCCTGGGCAACATGATG 2026

Qy       1067 GCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATG 1126
              | |    |     |   |   ||   | ||          || |    |
Db       2027 AAATGCCCTCTCTGCAAAAAAAAAAAAAAATTACAAAAATTGGCGGAGCATGGTGGTCCGT 2086

Qy       1127 GTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATT 1186
              | ||  |   ||  | | |    | ||  |    | |    ||  | | |||    |
Db       2087 GCCTGTGGTCCCAGCTACGCGGGAGGCTAAAGTGGGAGGATCGCTTGAGCCT-GGGAGGT 2145

Qy       1187 TCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTG 1246
              ||      |  |||||  |      |   |  | ||  ||  || | | | |
Db       2146 GAAGACTGCAGTGAGCTGTGATTGTACCACAGCCCTCTAGGCTGGGGGACAGACT----- 2200

Qy       1247 TAGACAGTAGTAGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAA 1306
              ||    || |   |  | | | | |  | ||  |   | ||| |
Db       2201 ----GAGACCCTGTTTCCCCTCCGCAAAAAAATTGACAAAAGTGTAATAAGAGGTGCCTG 2256

Qy       1307 ATATAGCCAGGTGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTG 1366
              |||| || ||| ||  |||||||||||||||||||||||||||||| ||| ||| | |
Db       2257 ATATGGCTAGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGG 2316

Qy       1367 TGGATCAC--GAGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTC 1424
              || ||||   |||||||| ||  |||||||||||||||||||| |||| ||| ||||
Db       2317 CGGGTCACCTAAGGTCAGGAGTGTGAGACCAGCCTGGCCAACATGGAGAAAGCCCATCTC 2376

Qy       1425 TACTAAAAAATACAAAAATTAGCCT--------GGTGTGGTGGCACGCATCTGTAATCCCA 1476
              | |||||||||||||| |   |          |  |||||||  |    ||||||||||
Db       2377 TTCTAAAAATACAAAATTAGCCGGCTGTGGGGGCAGTGGTGGAGCATGCCTGTAATCCCA 2436

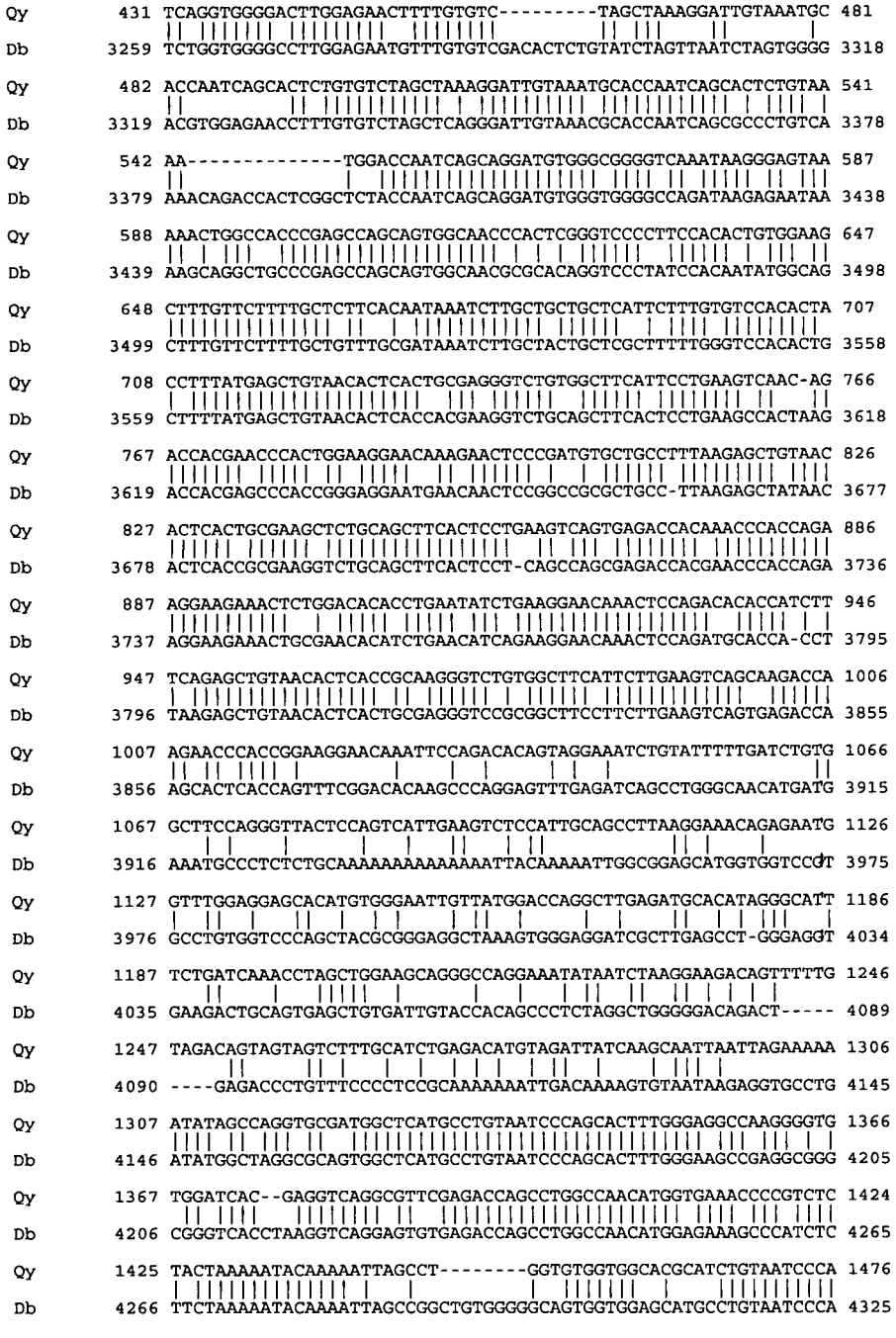
Qy       1477 G-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGA 1535
              | |||||||||||||||||||| ||||| |||||||   |||||| | ||||||||||
Db       2437 GCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGCGGTTGCAGTGA 2496

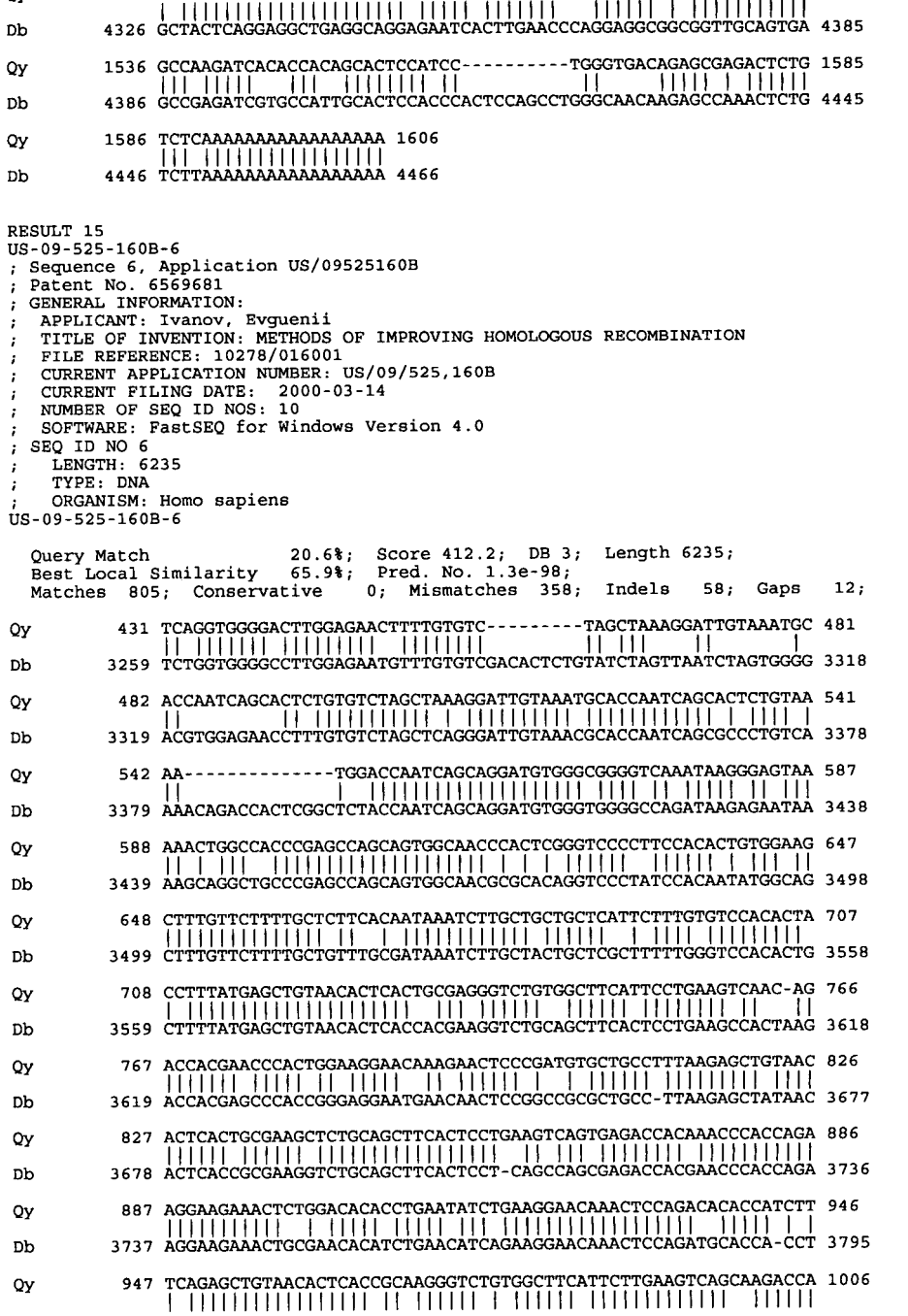
Qy       1536 GCCAAGATCACACCACAGCACTCCATCC----------TGGGTGACAGAGCGAGACTCTG 1585
              ||| |||||   |||  |||||||| ||          ||     ||||| | ||||||
Db       2497 GCCGAGATCGTGCCATTGCACTCCACCCACTCCAGCCTGGGCAACAAGAGCCAAACTCTG 2556

Qy       1586 TCTCAAAAAAAAAAAAAAAAA 1606
              ||| ||||||||||||||||
Db       2557 TCTTAAAAAAAAAAAAAAAAA 2577


RESULT 14
US-09-305-384-5
; Sequence 5, Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
;  APPLICANT: Treco, Douglas A.
;  APPLICANT: Heartlein, Michael W.
;  APPLICANT: Selden, Richard F
;  TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
;  FILE REFERENCE: 07236/017001
;  CURRENT APPLICATION NUMBER: US/09/305,384
;  CURRENT FILING DATE: 1999-05-05
;  EARLIER APPLICATION NUMBER: US 60/084,649
;  EARLIER FILING DATE: 1998-05-07
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 6235
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-305-384-5

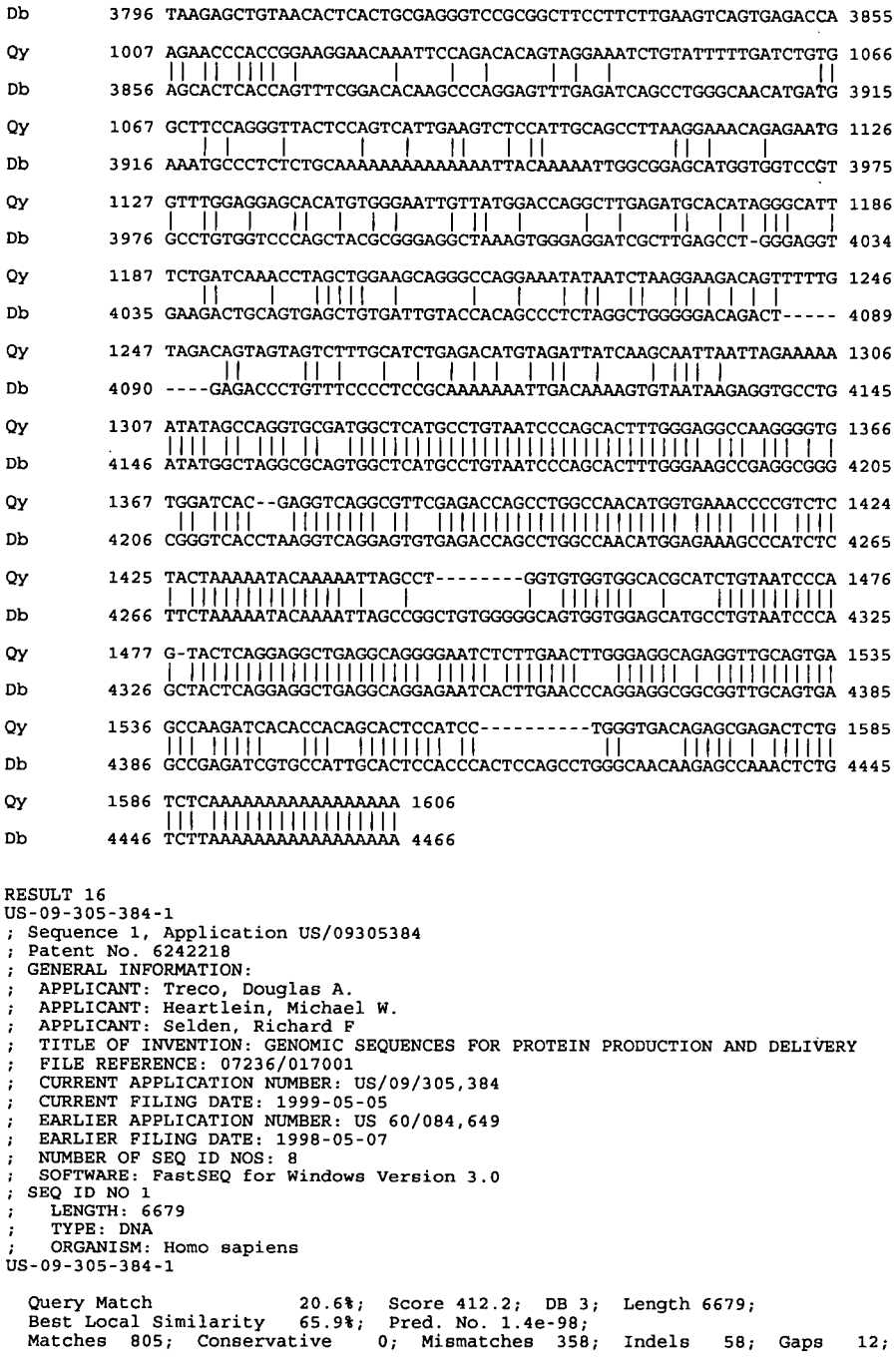
  Query Match             20.6%;  Score 412.2;  DB 3;  Length 6235;
  Best Local Similarity   65.9%;  Pred. No. 1.3e-98;
  Matches  805;  Conservative    0;  Mismatches  358;  Indels   58;  Gaps   12;
```





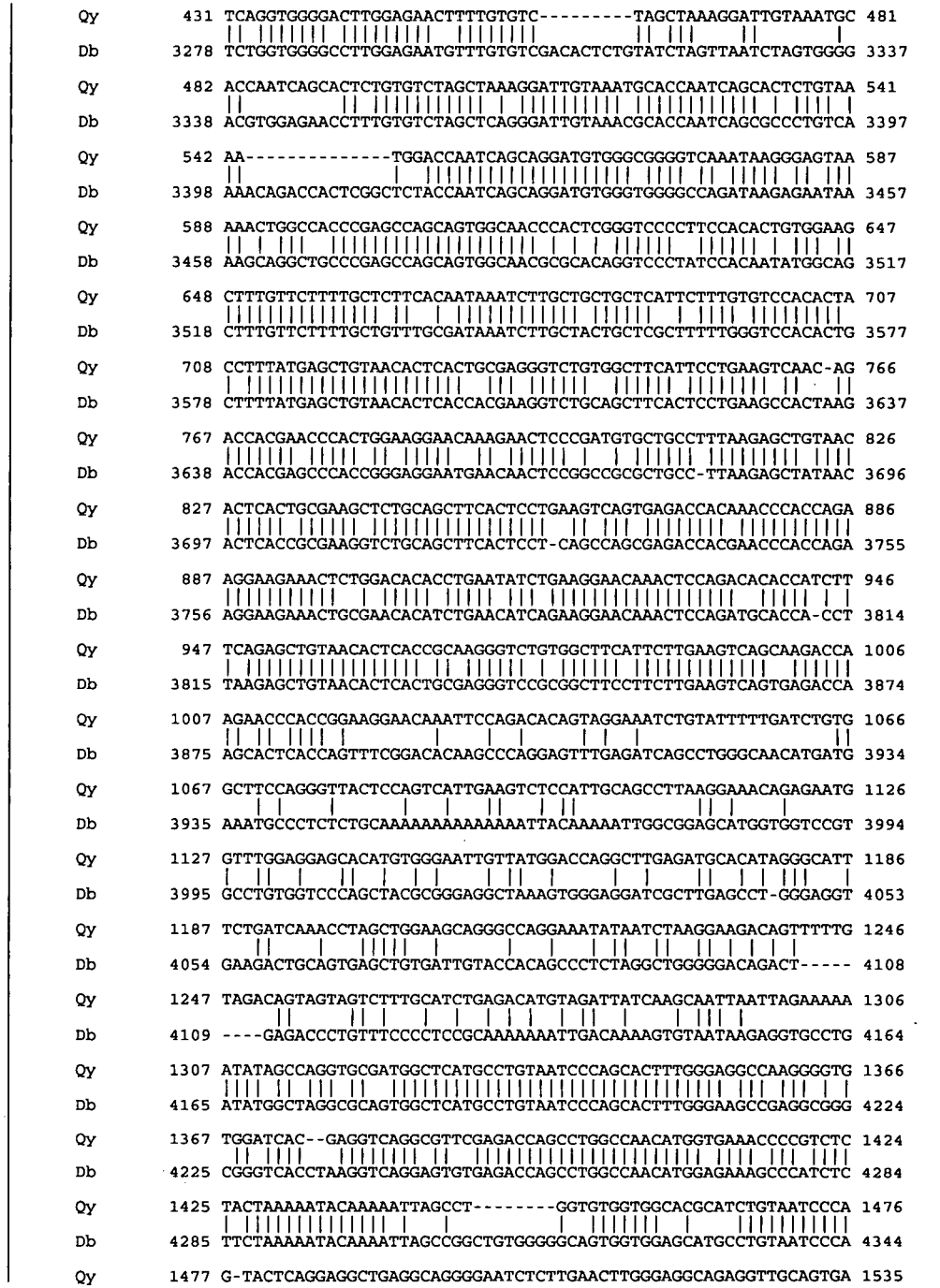
RESULT 15
US-09-525-160B-6
; Sequence 6, Application US/09525160B
; Patent No. 6569681
; GENERAL INFORMATION:
; APPLICANT: Ivanov, Evguenii
; TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION
; FILE REFERENCE: 10278/016001
; CURRENT APPLICATION NUMBER: US/09/525,160B
; CURRENT FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 6235
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-525-160B-6

Query Match 20.6%; Score 412.2; DB 3; Length 6235;
Best Local Similarity 65.9%; Pred. No. 1.3e-98;
Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;



RESULT 16
US-09-305-384-1
; Sequence 1, Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Selden, Richard F
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/017001
; CURRENT APPLICATION NUMBER: US/09/305,384
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: US 60/084,649
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 6679
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-384-1

Query Match 20.6%; Score 412.2; DB 3; Length 6679;
Best Local Similarity 65.9%; Pred. No. 1.4e-98;
Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;



```
Db       4345 GCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGCGGTTGCAGTGA 4404

Qy       1536 GCCAAGATCACACCACAGCACTCCATCC----------TGGGTGACAGAGCGAGACTCTG 1585

Db       4405 GCCGAGATCGTGCCATTGCACTCCACCCACTCCAGCCTGGGCAACAAGAGCCAAACTCTG 4464

Qy       1586 TCTCAAAAAAAAAAAAAAAAA 1606

Db       4465 TCTTAAAAAAAAAAAAAAAAA 4485


RESULT 17
US-09-525-160B-5
; Sequence 5, Application US/09525160B
; Patent No. 6569681
; GENERAL INFORMATION:
;  APPLICANT: Ivanov, Evguenii
;  TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION
;  FILE REFERENCE: 10278/016001
;  CURRENT APPLICATION NUMBER: US/09/525,160B
;  CURRENT FILING DATE:  2000-03-14
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 6679
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-525-160B-5

  Query Match             20.6%;  Score 412.2;  DB 3;  Length 6679;
  Best Local Similarity   65.9%;  Pred. No. 1.4e-98;
  Matches  805;  Conservative    0;  Mismatches  358;  Indels   58;  Gaps   12;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTC---------TAGCTAAAGGATTGTAAATGC 481

Db       3278 TCTGGTGGGGCCTTGGAGAATGTTTGTGTCGACACTCTGTATCTAGTTAATCTAGTGGGG 3337

Qy        482 ACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAA 541

Db       3338 ACGTGGAGAACCTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCGCCCTGTCA 3397

Qy        542 AA--------------TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAA 587

Db       3398 AAACAGACCACTCGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAA 3457

Qy        588 AAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAG 647

Db       3458 AAGCAGGCTGCCCGAGCCAGCAGTGGCAACGCGCACAGGTCCCTATCCACAATATGGCAG 3517

Qy        648 CTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTA 707

Db       3518 CTTTGTTCTTTTGCTGTTTGCGATAAATCTTGCTACTGCTCGCTTTTTGGGTCCACACTG 3577

Qy        708 CCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AG 766

Db       3578 CTTTTATGAGCTGTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCACTAAG 3637

Qy        767 ACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAAC 826

Db       3638 ACCACGAGCCCACCGGGAGGAATGAACAACTCCGGCCGCGCTGCC-TTAAGAGCTATAAC 3696

Qy        827 ACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGA 886

Db       3697 ACTCACCGCGAAGGTCTGCAGCTTCACTCCT-CAGCCAGCGAGACCACGAACCCACCAGA 3755

Qy        887 AGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTT 946

Db       3756 AGGAAGAAACTGCGAACACATCTGAACATCAGAAGGAACAAACTCCAGATGCACCA-CCT 3814

Qy        947 TCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCA 1006

Db       3815 TAAGAGCTGTAACACTCACTGCGAGGGTCCGCGGCTTCCTTCTTGAAGTCAGTGAGACCA 3874

Qy       1007 AGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTG 1066

Db       3875 AGCACTCACCAGTTTCGGACACAAGCCCAGGAGTTTGAGATCAGCCTGGGCAACATGATG 3934

Qy       1067 GCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATG 1126

Db       3935 AAATGCCCTCTCTGCAAAAAAAAAAAAAAATTACAAAAATTGGCGGAGCATGGTGGTCCGT 3994

Qy       1127 GTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATT 1186

Db       3995 GCCTGTGGTCCCAGCTACGCGGGAGGCTAAAGTGGGAGGATCGCTTGAGCCT-GGGAGGT 4053

Qy       1187 TCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTG 1246

Db       4054 GAAGACTGCAGTGAGCTGTGATTGTACCACAGCCCTCTAGGCTGGGGGACAGACT----- 4108

Qy       1247 TAGACAGTAGTAGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAA 1306

Db       4109 ----GAGACCCTGTTTCCCCTCCGCAAAAAAATTGACAAAAGTGTAATAAGAGGTGCCTG 4164

Qy       1307 ATATAGCCAGGTGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTG 1366

Db       4165 ATATGGCTAGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGG 4224

Qy       1367 TGGATCAC--GAGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTC 1424

Db       4225 CGGGTCACCTAAGGTCAGGAGTGTGAGACCAGCCTGGCCAACATGGAGAAAGCCCATCTC 4284

Qy       1425 TACTAAAAATACAAAAATTAGCCT--------GGTGTGGTGGCACGCATCTGTAATCCCA 1476

Db       4285 TTCTAAAAATACAAAATTAGCCGGCTGTGGGGGCAGTGGTGGAGCATGCCTGTAATCCCA 4344

Qy       1477 G-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGA 1535

Db       4345 GCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGCGGTTGCAGTGA 4404

Qy       1536 GCCAAGATCACACCACAGCACTCCATCC----------TGGGTGACAGAGCGAGACTCTG 1585

Db       4405 GCCGAGATCGTGCCATTGCACTCCACCCACTCCAGCCTGGGCAACAAGAGCCAAACTCTG 4464

Qy       1586 TCTCAAAAAAAAAAAAAAAAA 1606

Db       4465 TCTTAAAAAAAAAAAAAAAAA 4485


RESULT 18
US-09-949-016-12298
; Sequence 12298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12298
;   LENGTH: 115963
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(115963)
;   OTHER INFORMATION: n = A,T,C or G
```

US-09-949-016-12298

Query Match 20.6%; Score 411.8; DB 3; Length 115963;
Best Local Similarity 68.5%; Pred. No. 6.1e-98;
Matches 776; Conservative 0; Mismatches 282; Indels 75; Gaps 12;

```
Qy        493 CTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAA--------- 543
              || |||||||||| | |||||||||| |||||||||| | |||  |||
Db      74942 CTTTGTGTCTAGCTCAGGGATTGTAAACTCACCAATCAGCGCCCTGGCAAAACAGACCAC 75001

Qy        544 -----TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCAC 598
                  | |||||||| |||||||||| |||| || ||||| || ||||| | |||  |
Db      75002 TGGGCTCTACCAATCAACAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGC 75061

Qy        599 CCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTT 658
              | |||||||||||||||||| |||||||| |||||| ||||||||||||||||||||||
Db      75062 CGGAGCCAGCAGTGGCAACCCGCTCGGGTCTCCTTCCGCACTGTGGAAGCTTTGTTCTTT 75121

Qy        659 TGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGC 718
              |||                              | |||| |||| |||| | ||||||||||
Db      75122 TGC------------------------------TGTTTGGGTCCCCACTGCTTTTATGAGC 75152

Qy        719 TGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCC 777
              |||||||||||| |||| |||||| ||||||| |||||||| || | ||||||||| |||
Db      75153 TGTAACACTCACCGCGAAGGTCTGCAGCTTCACTCCTGAAGACAGCGAGACCACGAGCCC 75212

Qy        778 ACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGA 837
              ||  | |||||| || |||||| |||| ||  || ||||||||||||||||||| | ||
Db      75213 ACCAGGAGGAACGAACAACTCCAGATGCGCCACC-TTAAGAGCTGTAACACTCACCGTGA 75271

Qy        838 AGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACT 897
              || |||||||||||||||| ||| ||| |||||||| ||||||||||||||||||||||
Db      75272 AGGTCTGCAGCTTCACTCCT-AAGCCAGCGAGACCACGAACCCACCAGAAGGAAGAAACT 75330

Qy        898 CTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTA 957
              |   ||||| | ||| ||| ||| |||||||||||||| | ||| |||| |||||||||
Db      75331 CCAAACACATCCGAACATCAGAAAGAACAAACTCCAGACGCGCCA-CTTTAAGAGCTGTA 75389

Qy        958 ACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCG 1017
              |||||||||||||||||||| |||||||||||||||||| ||  ||||||||||| |||
Db      75390 ACACTCACCGCAAGGGTCTGCGGCTTCATTCTTGAAGTTAGTGAGACCAAGAACCTACC- 75448

Qy       1018 GAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGT 1077
              |||||| ||||||  ||||    |  |    | |  |    |          |
Db      75449 ---------AATTCCTGACACAAGAGGACCAATTAAGGCCTCAATTAGTGAGGGATTTGG 75499

Qy       1078 TACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAG 1137
              |   |          |||    ||      |  |    | ||    | | ||||
Db      75500 TGGCCTGAATTAGACAGTGGTTGTTAGGATGGAAGAGGTGGTGGGAGAATTCTAGAGGTA 75559

Qy       1138 CACATGTGGGAATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAAC 1197
                | |  | |      ||  ||    | | ||  | |   |  |      |
Db      75560 TTTAGGAAGTAGAATGGATTCACTGAACATTAGGGGAAGGAATAGTGGGGC-------AT 75612

Qy       1198 CTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGT 1257
                ||||| |      || | | |  || ||| |     |      ||||    |  || ||
Db      75613 AGGGCTGGCAAGTACTCCTGAAGAAGTATTCTGTGATCAAGGAAGTTTGGGAAATGTTGT 75672

Qy       1258 AGTCTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGG 1317
              | ||| |    |  ||      | | |||  ||  | | |    ||  ||  ||
Db      75673 ATTCTGTCATATTTGGA-------AGTCACAACACATATGTAAAAGCCCTTAGGGCTGGG 75725

Qy       1318 TGCGATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCAC--G 1375
              |  | ||||||| ||||||||| ||| |||||||||||||  |||   |||||||||  |
Db      75726 TATGGTGGCTCACGCCTGTAATTCCAACACTTTGGGAGGCTGAGGCAAGTGGATCACCTG 75785

Qy       1376 AGGTCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATA 1435
              |||||| | ||| |||| |||| |||||||||||||||| ||| | ||||||||||||||
Db      75786 AGGTCAAGAGTTTGAGAACAGCTTGGCCAACATGGTGAGACCTTGGCTCTACTAAAAATA 75845

Qy       1436 C-AAAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAG 1493
              | |||||||||| ||  |||| |||||   ||||| ||  || ||||||||||||||||
Db      75846 CAAAAAATTAGCTGGGCATGGTAGCACGTGCCTGTAGTCTAAGCTACTCAGGAGGCTGAG 75905
```

```
Qy       1494 GCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAG 1553
              | ||| |||||||||||||||  ||||||||| ||||||||||||||||||||   ||| |
Db      75906 GTAGGAGAATCTCTTGAACTCAGGAGGCAGATGTTGCAGTGAGCCAAGATCGTGCCATTG 75965

Qy       1554 CACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAA 1606
              ||||||| |||||| ||||||| |||||||||| |||||||||||||||||
Db      75966 CACTCCAGCCTGGGCGACAGAGTGAGACTCTGTCCAAAAAAAAAAAAAAAAAA 76018
```

RESULT 19
US-09-949-016-12707/c
; Sequence 12707, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12707
; LENGTH: 190078
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(190078)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12707

Query Match 20.3%; Score 407.2; DB 3; Length 190078;
Best Local Similarity 84.2%; Pred. No. 1.2e-96;
Matches 497; Conservative 0; Mismatches 83; Indels 10; Gaps 3;

```
Qy        437 GGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCT 496
              |||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||
Db      57447 GGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATCAGCACTCT 57388

Qy        497 GT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGAC 548
              ||          |   |   |   |  ||||| ||| ||||||| ||||||||||||||||
Db      57387 GTCAAAACGGACCACTCAGCACTCTGTAAAATGGACCGATCAGCTCTCTGTAAAATGGAC 57328

Qy        549 CAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACC-CGAGCCAG 607
              || ||||||  ||||||| |||| ||||||||||| ||||  | |||||||  |||||||
Db      57327 CAGTCAGCATTATGTGGGTGGGGCCAAATAAGGGAATAAAGGCAGGCCACCGGGAGCCAG 57268

Qy        608 CAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCA 667
              | | |||||||| ||| |||| ||||||||  |||||||||||||||||||| ||||||||
Db      57267 CTGCGGCAACCCGCTCTGGTCTCCTTCCACGGTGTGGAAGCTTTGTTCTTTCGCTCTTCA 57208

Qy        668 CAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACT 727
              |||||||||||||||||||| |||||| |||| |   ||||||||||||||| ||||
Db      57207 CAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCGCGTCACCTTTATGAGCTGTAGCACT 57148

Qy        728 CACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGG 786
              | |||| | ||||||  |||||| | |||| | || | |||||||  |||||  |||||
Db      57147 CCCTGCAAAGGTCTGCAGCTTCACTTCTGAGGCCAGCGAGACCACAAACCCACCAGAAGG 57088

Qy        787 AACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCA 846
              ||| || |||||| ||   |||||||||||||||| ||||||||| | |||| ||||||
Db      57087 AACGAACAACTCCAGACACGCTGCCTTTAAGAGCTGCAACACTCACCGTGAAGGTCTGCA 57028

Qy        847 GCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACA 906
              |||||| |||||||||||| |||||||| |||||||||||||||||||||||||||||||
```

```
Db      57027 GCTTCATTCCTGAAGTCAGCGAGACCACGAACCCACCAGAAGGAAGAAACTCTGGACACA 56968

Qy        907 CCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACC 966
              ||||| |||||||||||||||||  ||||||||||||| ||| || ||||||||||
Db      56967 TCTGAACATCTGAAGGAACAAACTCTGGACACACCATCTTTAAGAACTATAACACTCACT 56908

Qy        967 GCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016
              || ||||||||| ||||||||||||| || || ||||||||||||||||
Db      56907 GCGAGGGTCTGTCGCTTCATTCTTGAAATCGGCTAGACCAAGAACCCACC 56858


RESULT 20
US-09-949-016-17026/c
; Sequence 17026, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17026
;   LENGTH: 190078
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(190078)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17026

  Query Match             20.3%;  Score 407.2;  DB 3;  Length 190078;
  Best Local Similarity   84.2%;  Pred. No. 1.2e-96;
  Matches  497;  Conservative    0;  Mismatches   83;  Indels   10;  Gaps    3;

Qy        437 GGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCT 496
              ||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||
Db      57447 GGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATCAGCACTCT 57388

Qy        497 GT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGAC 548
              ||        |   |   |   |    ||||| ||| |||||| ||||||||||||||
Db      57387 GTCAAAACGGACCACTCAGCACTCTGTAAAATGGACCGATCAGCTCTCTGTAAAATGGAC 57328

Qy        549 CAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACC-CGAGCCAG 607
              || ||||||  ||||||| |||| |||||||||| ||||  | |||||||  |||||||
Db      57327 CAGTCAGCATTATGTGGGTGGGGCCAAATAAGGGAATAAAGGCAGGCCACCGGGAGCCAG 57268

Qy        608 CAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCA 667
              | | |||||||| ||| |||| |||||||| ||||||||||||||||||| ||||||||
Db      57267 CTGCGGCAACCCGCTCTGGTCTCCTTCCACGGTGTGGAAGCTTTGTTCTTTCGCTCTTCA 57208

Qy        668 CAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACT 727
              |||||||||||||||||||| |||||| |||| |   ||||||||||||||| ||||
Db      57207 CAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCGCGTCACCTTTATGAGCTGTAGCACT 57148

Qy        728 CACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGG 786
              | |||| | ||||||  |||||| | |||| | || | |||||||  |||||||  |||||
Db      57147 CCCTGCAAAGGTCTGCAGCTTCACTTCTGAGGCCAGCGAGACCACAAACCCACCAGAAGG 57088

Qy        787 AACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCA 846
              ||| || |||||| ||   |||||||||||||||| ||||||||| | |||| ||||||
Db      57087 AACGAACAACTCCAGACACGCTGCCTTTAAGAGCTGCAACACTCACCGTGAAGGTCTGCA 57028

Qy        847 GCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACA 906
              |||||| |||||||||||| |||||||| |||||||||||||||||||||||||||||
Db      57027 GCTTCATTCCTGAAGTCAGCGAGACCACGAACCCACCAGAAGGAAGAAACTCTGGACACA 56968

Qy        907 CCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACC 966
              ||||| |||||||||||||||||  ||||||||||||| ||| || ||||||||||
Db      56967 TCTGAACATCTGAAGGAACAAACTCTGGACACACCATCTTTAAGAACTATAACACTCACT 56908

Qy        967 GCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016
              || ||||||||| ||||||||||||| || || ||||||||||||||||
Db      56907 GCGAGGGTCTGTCGCTTCATTCTTGAAATCGGCTAGACCAAGAACCCACC 56858


RESULT 21
US-09-949-016-184449
; Sequence 184449, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 184449
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-184449

  Query Match             20.3%;  Score 406.8;  DB 3;  Length 601;
  Best Local Similarity   84.1%;  Pred. No. 1.3e-97;
  Matches  496;  Conservative    1;  Mismatches   83;  Indels   10;  Gaps    3;

Qy        437 GGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCT 496
              ||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||
Db          7 GGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATCAGCACTCT 66

Qy        497 GT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGAC 548
              ||        |   |   |   |    ||||| ||| |||||| ||||||||||||||
Db         67 GTCAAAACGGACCACTCAGCACTCTGTAAAATGGACCGATCAGCTCTCTGTAAAATGGAC 126

Qy        549 CAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACC-CGAGCCAG 607
              || ||||||  ||||||| |||| |||||||||| ||||  | |||||||  |||||||
Db        127 CAGTCAGCATTATGTGGGTGGGGCCAAATAAGGGAATAAAGGCAGGCCACCGGGAGCCAG 186

Qy        608 CAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCA 667
              | | |||||||| ||| |||| |||||||| ||||||||||||||||||| ||||||||
Db        187 CTGCGGCAACCCGCTCTGGTCTCCTTCCACGGTGTGGAAGCTTTGTTCTTTCGCTCTTCA 246

Qy        668 CAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACT 727
              |||||||||||||||||||| |||||| |||| |   ||||||||||||||: ||||
Db        247 CAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCGCGTCACCTTTATGAGCTGTRGCACT 306

Qy        728 CACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGG 786
              | |||| | ||||||  |||||| | |||| | || | |||||||  |||||||  |||||
Db        307 CCCTGCAAAGGTCTGCAGCTTCACTTCTGAGGCCAGCGAGACCACAAACCCACCAGAAGG 366

Qy        787 AACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCA 846
              ||| || |||||| ||   |||||||||||||||| ||||||||| | |||| ||||||
Db        367 AACGAACAACTCCAGACACGCTGCCTTTAAGAGCTGCAACACTCACCGTGAAGGTCTGCA 426

Qy        847 GCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACA 906
              |||||| |||||||||||| |||||||| |||||||||||||||||||||||||||||
Db        427 GCTTCATTCCTGAAGTCAGCGAGACCACGAACCCACCAGAAGGAAGAAACTCTGGACACA 486
```

```
Qy        907 CCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACC 966
              ||||| |||||||||||||||||  |||||||||||||| ||| || ||||||||||
Db        487 TCTGAACATCTGAAGGAACAAACTCTGGACACACCATCTTTAAGAACTATAACACTCACT 546

Qy        967 GCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016
              || ||||||||| |||||||||||||| || || ||||||||||||||||
Db        547 GCGAGGGTCTGTCGCTTCATTCTTGAAATCGGCTAGACCAAGAACCCACC 596


RESULT 22
US-09-949-016-14890/c
; Sequence 14890, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14890
;   LENGTH: 93364
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(93364)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14890

  Query Match             20.2%;  Score 404.8;  DB 3;  Length 93364;
  Best Local Similarity   67.6%;  Pred. No. 3.9e-96;
  Matches  719;  Conservative    0;  Mismatches  277;  Indels   68;  Gaps    8;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||||||||||||||||||||  |||||||||  ||||||||||  ||||||||||
Db      18468 TCTAGTGGGGACTTGGAGAACTTTTACGTCTAGCTAGAGGATTGTAAACACACCAATCAG 18409

Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCA---------------- 533
              |||||||||||||| | | ||||| ||||  ||||||||||
Db      18408 CACTCTGTGTCTAGATCAGGGATTATAAACACACCAATCAGCATCCTGTCAAAACAGACC 18349

Qy        534 --------CTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGT 585
                     |||||||||||| |||||||||||||||||||  ||| ||| ||||||||| |
Db      18348 AATCAGCTCTCTGTAAAATGGGCCAATCAGCAGGATGTGGCTGGGATCAGATAAGGGAAT 18289

Qy        586 AAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGA 645
              |||| | |||  ||||||||||||  |||||||| ||||||||||||||||||||||||
Db      18288 AAAAGCAGGCTGCCCGAGCCAGCATGGGCAACCCGCTCGGGTCCCCTTCCACACTGTGGA 18229

Qy        646 AGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACAC 705
              |||||||||||| ||||||  || ||||||||||||||||| ||||||  |||  |||
Db      18228 AGCTTTGTTCTTTCGCTCTTTGCAGTAAATCTTGCTGCTGCTCACTCTTTGGGTCTGCAC 18169

Qy        706 TACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC- 764
              | |||||||||||| |||||||||||| || ||||||  ||||||| |||||||| || |
Db      18168 TGCCTTTATGAGCTATAACACTCACTGAGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCG 18109

Qy        765 AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTG--------------- 809
              ||| |||||||||| ||  |||| ||| |||||  || | |  |
Db      18108 AGATCACGAACCCACCGGGGGGAATAAACAACTCTAGACGGGAGGAACGAACAACTCCAG 18049

Qy        810 -------CCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGT 862
                    | |||||||||||||||||||| | | || ||||||||||||||||||||
Db      18048 ACGTGCCACCTTAAGAGCTGTAACACTCACCGTGGAGGTCTGCAGCTTCACTCCTGAAGC 17989

Qy        863 CAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGG 922
              ||| |||||||| |||||||||||||||| |||||||| |||   | ||| ||| |||||
Db      17988 CAGCGAGACCACGAACCCACCAGAAGGAATAAACTCTGAACATGTCCGAACATCAGAAGG 17929

Qy        923 AACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCT 982
              || ||||||  ||||| | |||||| ||| |||||||||||||| |||||||| ||||
Db      17928 AAGAAACTCTGGACACGCTATCTTTAAGAACTGTAACACTCACCGCTAGGGTCTGCGGCT 17869

Qy        983 TCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTA 1042
              |||||||||||||||  |||||||||||||||           |||||| |||| | ||
Db      17868 TCATTCTTGAAGTCAGTGAGACCAAGAACCCACC----------AATTCCGGACATACTA 17819

Qy       1043 GGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATT 1102
              |  | || | |  |      |  |  ||  | | |  |  |     ||| || || |
Db      17818 GCTACTCAGAAGGCTGAGGTGGGAGGATC--GCCTGAGCCTGGGAGGCCAAGGCTGCAAT 17761

Qy       1103 GCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACATGTGGGAATTGTTATGGACCA 1162
              |   | | ||  | | | |    | |      | || | ||      |  | ||
Db      17760 GAGTCGATGAGTCATGATCGCA-CTACTATACTCCAGCCTGGGTGACAGAGCAATGCCCT 17702

Qy       1163 GGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTA-----GCTGGAAGCAGGGCCAG 1217
              | ||  |     | | |     ||| ||   |    ||    ||  | |         |
Db      17701 GTCTCCAAGGAAAAAAAAAAATGTTTGTGTGTATGTATATGTATGCGTGTATGTATCTGTG 17642

Qy       1218 GAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTTGCATCTGAGACAT 1277
              | |||      ||  || ||||       | |          || | |  | || | |
Db      17641 TATATACGTATAAAACAAAACAGACGCAAATAAGACAGCAGTACTGTCCCACAGACAAAG 17582

Qy       1278 GTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCGATGGCTCATGCCTGTA 1337
                ||     |||   | ||  | |||| | | |||||||   | ||||||||||||||
Db      17581 AATGAAACAGAAGAGTTAAAAAACAAAAGACACAGCCAGGCATGGTGGCTCATGCCTGTA 17522

Qy       1338 ATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGAT--CACGAGGTCAGGCGTTCGAGACCA 1395
              |||||||||||||||||| |   ||| | | |||| |  |||  |||| ||||||||||
Db      17521 ATCCCAGCACTTTGGGAAGTGGAGGTGGGAGGATTGCTTGAGTGCAGGAGTTCGAGACCA 17462

Qy       1396 GCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 1439
              |||| | |||||| |||| || || |||||| |    | | |||
Db      17461 GCCTAGGCAACATAGTGAGACACCATCTCTATTTCGTAAATAAA 17418


RESULT 23
US-09-949-016-12707
; Sequence 12707, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12707
;   LENGTH: 190078
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(190078)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12707

  Query Match             20.2%;  Score 403.8;  DB 3;  Length 190078;
  Best Local Similarity   84.8%;  Pred. No. 9.8e-96;
```

```
  Matches  513;  Conservative    0;  Mismatches   82;  Indels   10;  Gaps    5;

Qy        447 AGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT------GT 500
              || ||  |  |||||||||||||| |||||||||||||||||||||||||||||      |
Db       7506 AGCACCCTGTGGTCTAGCTAAAGGTTTGTAAATGCACCAATCAGCACTCTGTAAAACAGA 7565

Qy        501 CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGA 560
              | |   |      | |||||| || ||||||| |||||||||||||||| ||||||||||||
Db       7566 CCAATCAGCACTCTTTAAATGGACGAATCAGCGCTCTGTAAAATGGAGCAATCAGCAGGA 7625

Qy        561 TGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCA 620
              |||||  |||| |||||||||| |||||  |||||||||||||||||    |||||||||
Db       7626 CGTGGGTGGGGCCAAATAAGGGAATAAAAGCTGGCCACCCGAGCCAGC-CCGGCAACCCA 7684

Qy        621 CTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGC 680
              || |||||||||||||| |||||||||||||||||||  |||||||| ||||||||||
Db       7685 CTTGGGTCCCCTTCCACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAGTAAATCTTGC 7744

Qy        681 TGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTC 740
              |||||||| | |||| |||  |||||||||||||||||| ||||||||| || ||||||
Db       7745 TGCTGCTCACTGTTTGGGTCTGCACTACCTTTATGAGCTATAACACTCACCGC-AGGGTC 7803

Qy        741 TGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCC 799
              ||||| ||||||||||||||| | |||||||||||||| || || |||||| |||||
Db       7804 TGTGGTTTCATTCCTGAAGTCTGCGAGACCACGAACCCACCGGGAGAAACAAACAACTCT 7863

Qy        800 CGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGA 859
              ||||     |||||||||||| |||||||||||| |||| |||||  ||||||||||||
Db       7864 GGATGCCTCACCTTTAAGAGCTATAACACTCACTGTGAAGGTCTGCGACTTCACTCCTGA 7923

Qy        860 AGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGA 919
              | |||| |||||||| ||||||| |||||||||||||| ||| ||| ||||| ||||||
Db       7924 ACTCAGCGAGACCACGAACCCACTGGAAGGAAGAAACTCCGGATACATCTGAACATCTGA 7983

Qy        920 AGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTG 979
              |||||||||||||| |||||||||||||| ||| |||||||||||||| | |||||| |||
Db       7984 AGGAACAAACTCCGGACACACCATCTTTAAGAACTGTAACACTCACCACGAGGGTCCGTG 8043

Qy        980 GCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACA 1039
              |||||||||| ||||||||| |||||||||||||||| |||||||| |||||| ||||||
Db       8044 GCTTCATTCTAGAAGTCAGC-AGACCAAGAACCCACCAGAAGGAACCAATTCCGGACACA 8102

Qy       1040 GTAGG 1044
                |||
Db       8103 CCAGG 8107


RESULT 24
US-09-949-016-17026
; Sequence 17026, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17026
;   LENGTH: 190078
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(190078)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17026

  Query Match             20.2%;  Score 403.8;  DB 3;  Length 190078;
  Best Local Similarity   84.8%;  Pred. No. 9.8e-96;
  Matches  513;  Conservative    0;  Mismatches   82;  Indels   10;  Gaps    5;

Qy        447 AGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT------GT 500
              || ||  |  |||||||||||||| |||||||||||||||||||||||||||||      |
Db       7506 AGCACCCTGTGGTCTAGCTAAAGGTTTGTAAATGCACCAATCAGCACTCTGTAAAACAGA 7565

Qy        501 CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGA 560
              | |   |      | |||||| || ||||||| |||||||||||||||| ||||||||||||
Db       7566 CCAATCAGCACTCTTTAAATGGACGAATCAGCGCTCTGTAAAATGGAGCAATCAGCAGGA 7625

Qy        561 TGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCA 620
              |||||  |||| |||||||||| |||||  |||||||||||||||||    |||||||||
Db       7626 CGTGGGTGGGGCCAAATAAGGGAATAAAAGCTGGCCACCCGAGCCAGC-CCGGCAACCCA 7684

Qy        621 CTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGC 680
              || |||||||||||||| |||||||||||||||||||  |||||||| ||||||||||
Db       7685 CTTGGGTCCCCTTCCACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAGTAAATCTTGC 7744

Qy        681 TGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTC 740
              |||||||| | |||| |||  |||||||||||||||||| ||||||||| || ||||||
Db       7745 TGCTGCTCACTGTTTGGGTCTGCACTACCTTTATGAGCTATAACACTCACCGC-AGGGTC 7803

Qy        741 TGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCC 799
              ||||| ||||||||||||||| | |||||||||||||| || || |||||| |||||
Db       7804 TGTGGTTTCATTCCTGAAGTCTGCGAGACCACGAACCCACCGGGAGAAACAAACAACTCT 7863

Qy        800 CGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGA 859
              ||||     |||||||||||| |||||||||||| |||| |||||  ||||||||||||
Db       7864 GGATGCCTCACCTTTAAGAGCTATAACACTCACTGTGAAGGTCTGCGACTTCACTCCTGA 7923

Qy        860 AGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGA 919
              | |||| |||||||| ||||||| |||||||||||||| ||| ||| ||||| ||||||
Db       7924 ACTCAGCGAGACCACGAACCCACTGGAAGGAAGAAACTCCGGATACATCTGAACATCTGA 7983

Qy        920 AGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTG 979
              |||||||||||||| |||||||||||||| ||| |||||||||||||| | |||||| |||
Db       7984 AGGAACAAACTCCGGACACACCATCTTTAAGAACTGTAACACTCACCACGAGGGTCCGTG 8043

Qy        980 GCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACA 1039
              |||||||||| ||||||||| |||||||||||||||| |||||||| |||||| ||||||
Db       8044 GCTTCATTCTAGAAGTCAGC-AGACCAAGAACCCACCAGAAGGAACCAATTCCGGACACA 8102

Qy       1040 GTAGG 1044
                |||
Db       8103 CCAGG 8107


RESULT 25
US-10-104-047-463
; Sequence 463, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. 6943241el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 463
;   LENGTH: 2942
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-104-047-463

  Query Match             20.0%;  Score 401.2;  DB 3;  Length 2942;
```

Best Local Similarity 83.7%; Pred. No. 7.6e-96;
Matches 497; Conservative 0; Mismatches 78; Indels 19; Gaps 3;

```
Qy        453 TTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGA 512
Db       2116 TCTGTGTCTAGCTGAAGGTTTGTAAATGCACCAATCAGCACTCCGTA------------ 2162

Qy        513 TTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGT 572
Db       2163 ----AAACGGACTGATCAGTGCTTTGTAAAATGGACCAATCAGCAGGATGTAGGCAGGGC 2218

Qy        573 CAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCT 632
Db       2219 CAAATAAGGGAAAAAAAGCTGG-CACCCAAGCCAACAGGGGCAACCTGCTTGAGTCCCTT 2277

Qy        633 TCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTC 692
Db       2278 TCCACATTGTGGAAGCTTTGTTCCTTCACTCTTCATGATAAATCTTGCTGCTGCTCACTC 2337

Qy        693 TTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATT 752
Db       2338 TTTGGGTCCGCACTACCTTTATGAGCTGTAACACTCACCACGAGGGTCTGTGGCTTCATT 2397

Qy        753 CCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCC 811
Db       2398 CCTTAAGTTAGCAAGACCACGAACCCACAGGGAGGAACAAACAACTCCGGGCACACCACC 2457

Qy        812 TTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGAC 871
Db       2458 TTTAAGAGCTATAACACTCACTGCGAAGGTCTGTGGCTTCACTCCTGTAGTCAGCAAAAC 2517

Qy        872 CACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTC 931
Db       2518 CACAAACCCACCGGAAGGAAGAAACTCCAGACACATCTAAACATCTGAAGGAACAAACTC 2577

Qy        932 CAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTG 991
Db       2578 TGGACACACCATCTTTAAGAACTGTAACACTCACCGTGAGGGTCCGTGGCTTCATTCTTG 2637

Qy        992 AAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
Db       2638 AAGTCAGGGAGACCAAGAACCCACCGGAAGGAACAAACTCCGGACACACTAGCA 2691
```

RESULT 26
US-09-949-016-15532/c
; Sequence 15532, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15532
; LENGTH: 107421
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15532

Query Match 19.6%; Score 392.2; DB 3; Length 107421;
Best Local Similarity 78.8%; Pred. No. 8.9e-93;
Matches 524; Conservative 0; Mismatches 98; Indels 43; Gaps 3;

```
Qy        419 TGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAA 478
Db      68426 TGTGTAGCTAATTGGGTAGGGGACTTGGAGACCTTTTGTGTCTAGCTAAAGGTTTGTAAA 68367

Qy        479 TGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGT---------------------- 516
Db      68366 TGCAGCAATCAGCACTCTGTCAAAACGGACCAATCAGTGCTATGTAAAATGGACCAATCA 68307

Qy        517 ----------AAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGG 566
Db      68306 GCCCCCTGTAAAACGGACCAATCAGCTCTCTGTAAAATGGACCAATCTGCAGGATGTGGG 68247

Qy        567 CGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGG 626
Db      68246 TGGGGTCAGATAAGGGAATAAAAGCAGGCTGCCTAAGCCAGCAGTGGCAACCCGCTCAGG 68187

Qy        627 TCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGC 686
Db      68186 TACCCTTCCACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAATAAATCTTGCTGCTGC 68127

Qy        687 TCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGC 746
Db      68126 TCACTCTTTGGGTCCGCGCCACCTTTATGAGCTGTAACACTCACCGCGAAGGTCTGCAGC 68067

Qy        747 TTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGT 805
Db      68066 TTCACTCCTGAAGCCAGCAAGACCACAAACCCACCAGAAGGAACGAACAACTCCAGACGC 68007

Qy        806 GCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAG 865
Db      68006 GCTGCCTTTAAGAGCTGTAACACTCATCGCAAAGGTCTGCAGCTTCACTCCTGAAGACAG 67947

Qy        866 TGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAAC 925
Db      67946 CAAGACTACGAACCCACCAGAAGGAAGAAACTCCAGACACAGCTGAACATCTGAAGGAAC 67887

Qy        926 AAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCA 985
Db      67886 AAACTCCGGACATACCATCTTTAAGAACCGTAACACTCATCGCGAGGGTCCACAGCTTCA 67827

Qy        986 TTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
Db      67826 TTCTTGAAGTTGGTGAGACCAAAAACCCACC----------AATTCCGGACACAGTAGGA 67777

Qy       1046 AATCT 1050
Db      67776 AAGCT 67772
```

RESULT 27
US-09-949-016-12735/c
; Sequence 12735, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12735
; LENGTH: 80355
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12735

```
  Query Match             19.1%;  Score 382;  DB 3;  Length 80355;
  Best Local Similarity   66.4%;  Pred. No. 3.9e-90;
  Matches  785;  Conservative    0;  Mismatches  210;  Indels  187;  Gaps    9;

Qy        437 GGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCT 496
              ||||||||||||||||||||||||||||| | |||||||||| ||||||||||||||| ||
Db       8672 GGGGACTTGGAGAACTTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCACCCT 8613

Qy        497 GT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGAC 548
              ||          | | |       |   |||   |||||| || ||||||||||||||||
Db       8612 GTCAAAACGGACCAATCAGCTCTCTGTAAAACAGACCAATTGGCTCTCTGTAAAATGGAC 8553

Qy        549 CAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGC 608
              |||||||||||||||||| || | || ||||| || ||||| | |||  ||| |||||||
Db       8552 CAATCAGCAGGATGTGGGTGGTGCCAGATAAGAGAATAAAAGCAGGCTGCCCAAGCCAGC 8493

Qy        609 AGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCAC 668
              |||||||||   |||  |||||||||||| ||||||||||||||||||  | ||| |
Db       8492 AGTGGCAACATGCTCCAGTCCCCTTCCACAATGTGGAAGCTTTGTTCTTTCACCCTTTAA 8433

Qy        669 AATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTC 728
              |||||||||||||||| |||| |||||| |||||  || | ||||||||||| |||||||
Db       8432 AATAAATCTTGCTGCTACTCACTCTTTGGGTCCAGGCTGCTTTTATGAGCTGCAACACTC 8373

Qy        729 ACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGA 787
              ||||| | ||||||| |||||  |||||| | || | ||||||||||||||  | ||||
Db       8372 ACTGCAAAGGTCTGTAGCTTCTCTCCTGAGGCCAGCAAGACCACGAACCCACCAGGAGGA 8313

Qy        788 ACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAG 847
              |  || || ||| |||| || ||| || |||||  |||||||||| |||||| |||||||
Db       8312 ATGAACAAGTCCAGATGCGCCGCC-TTCAGAGCGATAACACTCACCGCGAAGGTCTGCAG 8254

Qy        848 CTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACAC 907
              ||||||||||||| |||||||||||| |||||||||||||||||||||||||  |||||
Db       8253 CTTCACTCCTGAAGCCAGTGAGACCACGAACCCACCAGAAGGAAGAAACTCTCAACACAT 8194

Qy        908 CTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCG 967
              |||||  || |||||||||||||||||||||||    ||| ||| ||||||||||||||
Db       8193 CTGAACTTCAGAAGGAACAAACTCCAGACACACTGCTTTTAAGAACTGTAACACTCACCG 8134

Qy        968 CAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAA 1027
              ||||||||  || ||  ||||||||||||  ||||||||||||||          |
Db       8133 TGAGGGTCTGAGGTTTAGTTCTTGAAGTCAGTGAGACCAAGAACCCACC----------A 8084

Qy       1028 ATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACTCCAGTC 1087
              | | | ||||||  || |  | | |       ||   | |  ||
Db       8083 ACTGCGGACACACCAGCATTTAAGAAAGACAAATTACTTGACTC---------------- 8040

Qy       1088 ATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACATGTGGG 1147
                | ||||| || |    ||  ||| | ||
Db       8039 -------TTTCCATGGCGGAAAAAAAAAAAGACAG------------------------ 8012

Qy       1148 AATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCTGGAA 1207
Db       8011 ------------------------------------------------------------ 8012

Qy       1208 GCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTTGCA 1267
                                                |  |    |  | ||  || ||
Db       8011 -----------------------------------TTTAAACCTCATTGTAATTACA 7990

Qy       1268 TCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCGATGGCT 1327
                | || || | || |  ||  |                 |||  |  ||  |||||
Db       7989 ACAGATACTTTTAAAAATTCTCTC----------------TGGGCCGGACGCAGTGGCT 7947

Qy       1328 CATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCAC--GAGGTCAGGCG 1385
              || |||||||||||||||||||||||||||| ||| | |  ||||||  |||||||| |
Db       7946 CACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAG 7887

Qy       1386 TTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAG 1445
              |||||||||| |||||||||||||||||||||||||| ||| |||||||| ||||||||||
Db       7886 TTCGAGACCAACCTGGCCAACATGGTGAAACCCCATCTTTACTAAAACTACAAAAATTAA 7827

Qy       1446 CCTGGTGTGGTGGCACGCATCTGTAA--TCCCAGTACTCAGGAGGCTGAGGCAGGGGAAT 1503
              || ||||||||||   ||| || |||   ||   ||||| ||||||||| ||| | ||||
Db       7826 CCAGGTGTGGTGGTTTGCACCTATAAATCCCAGCTACTCGGGAGGCTGATGCATGAGAAT 7767

Qy       1504 CTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCATCC 1563
                |||||||   |||||||||||||||||| |||||||||||| ||||| |||||||| ||
Db       7766 TGCTTGAACCCTGGAGGCAGAGGTTGCAGCGAGCCAAGATCATACCACTGCACTCCAGCC 7707

Qy       1564 TGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAA 1605
              ||||  ||||||  ||||||    |   |||||||||| |||||
Db       7706 TGGGCAACAGAGTGAGACTACCTACAAAAAAAAAAAGAAAAA 7665


RESULT 28
US-09-949-016-13572/c
; Sequence 13572, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13572
;   LENGTH: 80357
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13572

  Query Match             19.1%;  Score 382;  DB 3;  Length 80357;
  Best Local Similarity   66.4%;  Pred. No. 3.9e-90;
  Matches  785;  Conservative    0;  Mismatches  210;  Indels  187;  Gaps    9;

Qy        437 GGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCT 496
              ||||||||||||||||||||||||||||| | |||||||||| ||||||||||||||| ||
Db       8672 GGGGACTTGGAGAACTTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCACCCT 8613

Qy        497 GT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGAC 548
              ||          | | |       |   |||   |||||| || ||||||||||||||||
Db       8612 GTCAAAACGGACCAATCAGCTCTCTGTAAAACAGACCAATTGGCTCTCTGTAAAATGGAC 8553

Qy        549 CAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGC 608
              |||||||||||||||||| || | || ||||| || ||||| | |||  ||| |||||||
Db       8552 CAATCAGCAGGATGTGGGTGGTGCCAGATAAGAGAATAAAAGCAGGCTGCCCAAGCCAGC 8493

Qy        609 AGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCAC 668
              |||||||||   |||  |||||||||||| ||||||||||||||||||  | ||| |
Db       8492 AGTGGCAACATGCTCCAGTCCCCTTCCACAATGTGGAAGCTTTGTTCTTTCACCCTTTAA 8433

Qy        669 AATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTC 728
              |||||||||||||||| |||| |||||| |||||  || | ||||||||||| |||||||
Db       8432 AATAAATCTTGCTGCTACTCACTCTTTGGGTCCAGGCTGCTTTTATGAGCTGCAACACTC 8373

Qy        729 ACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGA 787
              ||||| | ||||||| |||||  |||||| | || | ||||||||||||||  | ||||
Db       8372 ACTGCAAAGGTCTGTAGCTTCTCTCCTGAGGCCAGCAAGACCACGAACCCACCAGGAGGA 8313

Qy        788 ACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAG 847
              |  || || ||| |||| || ||| || |||||  |||||||||| |||||| |||||||
Db       8312 ATGAACAAGTCCAGATGCGCCGCC-TTCAGAGCGATAACACTCACCGCGAAGGTCTGCAG 8254

Qy        848 CTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACAC 907
              ||||||||||||| |||||||||||| |||||||||||||||||||||||||  |||||
Db       8253 CTTCACTCCTGAAGCCAGTGAGACCACGAACCCACCAGAAGGAAGAAACTCTCAACACAT 8194
```

```
Qy        908 CTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCG 967
              |||||  || ||||||||||||||||||||||||    ||| ||| ||||||||||||||||
Db       8193 CTGAACTTCAGAAGGAACAAACTCCAGACACACTGCTTTTAAGAACTGTAACACTCACCG 8134

Qy        968 CAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAACAA 1027
              ||||||||  || ||  |||||||||||||  ||||||||||||||||           |
Db       8133 TGAGGGTCTGAGGTTTAGTTCTTGAAGTCAGTGAGACCAAGAACCCACC----------A 8084

Qy       1028 ATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGGTTACTCCAGTC 1087
              | | | ||||||  || |   | | |       ||    | |  ||
Db       8083 ACTGCGGACACACCAGCATTTAAGAAAGACAAATTACTTGACTC---------------- 8040

Qy       1088 ATTGAAGTCTCCATTGCAGCCTTAAGGAAACAGAGAATGGTTTGGAGGAGCACATGTGGG 1147
                | ||||| || |    ||  ||| | ||
Db       8039 -------TTTCCATGGCGGAAAAAAAAAAAAGACAG------------------------ 8012

Qy       1148 AATTGTTATGGACCAGGCTTGAGATGCACATAGGGCATTTCTGATCAAACCTAGCTGGAA 1207

Db       8011 ------------------------------------------------------------ 8012

Qy       1208 GCAGGGCCAGGAAATATAATCTAAGGAAGACAGTTTTTGTAGACAGTAGTAGTCTTTGCA 1267
                                                      | |   |  | ||  || ||
Db       8011 -------------------------------------TTTAAACCTCATTGTAATTACA 7990

Qy       1268 TCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGGTGCGATGGCT 1327
              | || || | || |   ||   |                  ||| |  ||  |||||
Db       7989 ACAGATACTTTTAAAAATTCTCTC----------------TGGGCCGGACGCAGTGGCT 7947

Qy       1328 CATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGGGTGTGGATCAC--GAGGTCAGGCG 1385
              || |||||||||||||||||||||||||| ||| | |  ||||||  |||||||| |
Db       7946 CACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAG 7887

Qy       1386 TTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAG 1445
              |||||||||| |||||||||||||||||||||||| ||| ||||||| ||||||||||
Db       7886 TTCGAGACCAACCTGGCCAACATGGTGAAACCCCATCTTTACTAAAACTACAAAAATTAA 7827

Qy       1446 CCTGGTGTGGTGGCACGCATCTGTAA--TCCCAGTACTCAGGAGGCTGAGGCAGGGGAAT 1503
              || ||||||||||   ||| || |||   ||   ||||| ||||||||| ||| | ||||
Db       7826 CCAGGTGTGGTGGTTTGCACCTATAAATCCCAGCTACTCGGGAGGCTGATGCATGAGAAT 7767

Qy       1504 CTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCATCC 1563
                |||||||   |||||||||||||||||| |||||||||||| ||||| |||||||| ||
Db       7766 TGCTTGAACCCTGGAGGCAGAGGTTGCAGCGAGCCAAGATCATACCACTGCACTCCAGCC 7707

Qy       1564 TGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAA 1605
              ||||  |||||| ||||||    |   ||||||||||| |||||
Db       7706 TGGGCAACAGAGTGAGACTACCTACAAAAAAAAAAAGAAAAA 7665
```

RESULT 29
US-09-949-016-184415/c
; Sequence 184415, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 184415
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human

US-09-949-016-184415

Query Match 19.0%; Score 379.4; DB 3; Length 601;
Best Local Similarity 84.5%; Pred. No. 2.2e-90;
Matches 474; Conservative 1; Mismatches 77; Indels 9; Gaps 4;

```
Qy        447 AGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGT------GT 500
              || ||  |  |||||||||||||| ||||||||||||||||||||||||||      |
Db        570 AGCACCCTGTGGTCTAGCTAAAGGTTTGTAAATGCACCAATCAGCACTCTGTAAAACAGA 511

Qy        501 CTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGA 560
              | |   |      | |||||| || ||||||| |||||||||||||| ||||||||||||
Db        510 CCAATCAGCACTCTTTAAATGGACGAATCAGCGCTCTGTAAAATGGAGCAATCAGCAGGA 451

Qy        561 TGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCA 620
              ||||| |||| |||||||||| ||||| ||||||||||||||||   |||||||| ||||||||||
Db        450 CGTGGGTGGGGCCAAATAAGGGAATAAAAGCTGGCCACCCGAGCCAGC-CCGGCAACCCA 392

Qy        621 CTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGC 680
              || |||||||||||||| ||||||||||||||||||||   |||||||| ||||||||||
Db        391 CTTGGGTCCCCTTCCACGCTGTGGAAGCTTTGTTCTTTCACTCTTCACAGTAAATCTTGC 332

Qy        681 TGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTC 740
              ||||||||| | |||| |||  |||||||:|||||||| |||||||||| || ||||||
Db        331 TGCTGCTCACTGTTTGGGTCTGCACTACCTKTATGAGCTATAACACTCACCGC-AGGGTC 273

Qy        741 TGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCC 799
              ||||| ||||||||||||||||  | |||||||||||||| || || |||||| |||||
Db        272 TGTGGTTTCATTCCTGAAGTCTGCGAGACCACGAACCCACCGGGAGAAACAAACAACTCT 213

Qy        800 CGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGA 859
              ||||     |||||||||||| |||||||||| |||| |||||  ||||||||||||
Db        212 GGATGCCTCACCTTTAAGAGCTATAACACTCACTGTGAAGGTCTGCGACTTCACTCCTGA 153

Qy        860 AGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGA 919
              | |||| |||||||| |||||||  |||||||||||||| ||| ||| ||||| ||||||
Db        152 ACTCAGCGAGACCACGAACCCACTGGAAGGAAGAAACTCCGGATACATCTGAACATCTGA 93

Qy        920 AGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTG 979
              |||||||||||||| |||||||||||||| ||| ||||||||||||||| | |||||| |||
Db         92 AGGAACAAACTCCGGACACACCATCTTTAAGAACTGTAACACTCACCACGAGGGTCCGTG 33

Qy        980 GCTTCATTCTTGAAGTCAGCA 1000
              |||||||||| ||||||||||
Db         32 GCTTCATTCTAGAAGTCAGCA 12
```

RESULT 30
US-09-949-016-12805/c
; Sequence 12805, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12805
; LENGTH: 421491
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(421491)

```
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12805

  Query Match             18.6%;  Score 372;  DB 3;  Length 421491;
  Best Local Similarity   64.8%;  Pred. No. 3.5e-87;
  Matches  781;  Conservative    0;  Mismatches  340;  Indels   85;  Gaps   12;

Qy        494 TCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAA---------- 543
              | |||||||||||| | |||||||||| | |||||||||| | |||| |||
Db      16909 TTTGTGTCTAGCTCAGGGATTGTAAACGTACCAATCAGCGCCCTGTCAAAACAGACGACT 16850

Qy        544 ----TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACC 599
                  || |||||||||||||||||||| |||| || ||||| || ||||| | |||  |
Db      16849 GGGCTGTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCA 16790

Qy        600 CGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTT 659
              | |||||||| |||||||| ||| |||||| || |||| ||||||||||||||||||||
Db      16789 GGCGCCAGCAGCGGCAACCCTCTCCGGTCCCTTTGCACATTGTGGAAGCTTTGTTCTTTC 16730

Qy        660 GCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCT 719
              ||| ||  ||| ||| |||||||||||||| |||||| ||||||||| ||||||||||||
Db      16729 GCTGTTTGCAACAAACCTTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCT 16670

Qy        720 GTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCA 778
              |||||| ||||  ||| ||||||  |||||| |||||||| |||| ||||||||||| |
Db      16669 GTAACAGTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCAACGAGACCACGAACCTA 16610

Qy        779 CTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAA 838
              |  | |||||| ||  ||||| || | || ||| ||||||||||||||||||| |||||
Db      16609 CAAGGAGGAACGAACGACTCCAGACGCGCCGCC-TTAAGAGCTGTAACACTCACCGCGAA 16551

Qy        839 GCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTC 898
              | || |||||||||||| || || ||| |||||||| ||||||| |||| ||||||||||
Db      16550 GGTCCGCAGCTTCACTCTTG-AGCCAGCGAGACCACGAACCCACTAGAAAGAAGAAACTC 16492

Qy        899 TGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAA 958
              | ||||| | ||| ||| ||||||||||||||||||||| ||  |||| ||| ||||||
Db      16491 CGAACACATCCGAACATCAGAAGGAACAAACTCCAGACACGCCGCCTTTAAGAACTGTAA 16432

Qy        959 CACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC-- 1016
              |||||||||| |||||| | |||||||||| |||||||||| ||||||||||||||||
Db      16431 CACTCACCGCGAGGGTCCGCGGCTTCATTTTTGAAGTCAGTGAGACCAAGAACCCACCAA 16372

Qy       1017 ----------------------GGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTA 1053
                                   |  || | |  |  |  ||| |  |  | |
Db      16371 TTCCGGACACACTCAGAGCCCAAGATAGCATCCTACCCAGCACAAGAATAGCTCACCTTG 16312

Qy       1054 TTTTTGATCTGTGGCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAG 1113
              || | ||   ||| |   |    |  | |  ||| | |      | |
Db      16311 ATTGTCATTGCTGGTTGGTATTTAATGGGGAACCTTGACAGTTCCAACCCTTTTCCTTGA 16252

Qy       1114 GA---AACAGAGAATGGTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGCTTGAG 1170
              |    | ||| ||| |   | |   ||  ||| | | ||       |      |
Db      16251 GGCCTATCAGTGAAAGCAGTTGTCCAGTTCATCTTGTAAGACCAGCAGGGTCTCCAGAGC 16192

Qy       1171 ATGCACATAGGGCA--------TTTCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAAT 1222
              ||| ||| |   ||        | ||   |       |||||||   | | ||  | |
Db      16191 ATGGACAGACTACAGGGTAGGTTATCATTTGTCCTAAGGCTGGAAAATGTGGCATCACAG 16132

Qy       1223 ATAATCTAAGGAAGACAGTTTTTGTAGA--CAGTAGTAGTCTTTGCATCTGAGACATGTA 1280
                  |   ||| |||  || | |||  |  || |||||    |  |  |||   |
Db      16131 GATTCATCCAAAAGGCAGCATTGGGAGAGTCTATACTAGTCAGCACTCCAAAGAGTGTAA 16072

Qy       1281 GATTATCAAGCAATTAATTAGAAAAAATATA-------------------GCCAGGTGC 1320
              |   | ||  |  | |       |   | |   |                   ||  | |
Db      16071 GTCAACCAGGAGAATCCAGGGGTCAGACCAAGGATCAAAAACCAGGGAGGGGCACAGCGT 16012

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGG--GGTGTGGATCACGAGG 1378
              | |||||||||||| |||||||||| |||||||| |  |||  || |    ||  |||
Db      16011 GGTGGCTCATGCCTATAATCCCAGCCCTTTGGGAGACAGAGGCAGGAGGTTTTCTTGAGC 15952

Qy       1379 TCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACA- 1437
                |||  ||| |||||||||| ||  ||||| | || | || |||||||| ||||||| |

Db      15951 CCAGCAGTTTGAGACCAGCCCGGGTAACATAGGGAGATCCTGTCTCTACAAAAAATAAAT 15892

Qy       1438 ------AAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCT 1490
                    |||| ||||| ||  | ||| ||| |  || || |||||| ||||      ||
Db      15891 AAAAATAAAAATAGCCAGGCATTGTGACACACTCCTATAGTCCCAGCTACTTGCAGAGCA 15832

Qy       1491 GAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCA 1550
              |||   || | |||  |||| |   | | | ||||| ||||||||||   |||||  |||
Db      15831 GAGATGGGAGGATCATTTGAGCCCAGCAAGTAGAGGCTGCAGTGAGCTGTGATCATGCCA 15772

Qy       1551 CAGCACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAGG 1610
              | |||||||| | ||||  |||||| |||||  |||||||||||||||||||||||||
Db      15771 CTGCACTCCAGCATGGGCAACAGAGGGAGACCTTGTCTCAAAAAAAAAAAAAAAAAATAC 15712

Qy       1611 AAAGGA 1616
                |||||
Db      15711 CAAGGA 15706


RESULT 31
US-09-949-016-14060/c
; Sequence 14060, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14060
;   LENGTH: 421494
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(421494)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14060

  Query Match             18.6%;  Score 372;  DB 3;  Length 421494;
  Best Local Similarity   64.8%;  Pred. No. 3.5e-87;
  Matches  781;  Conservative    0;  Mismatches  340;  Indels   85;  Gaps   12;

Qy        494 TCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAA---------- 543
              | |||||||||||| | |||||||||| | |||||||||| | |||| |||
Db      16909 TTTGTGTCTAGCTCAGGGATTGTAAACGTACCAATCAGCGCCCTGTCAAAACAGACGACT 16850

Qy        544 ----TGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACC 599
                  || |||||||||||||||||||| |||| || ||||| || ||||| | |||  |
Db      16849 GGGCTGTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCA 16790

Qy        600 CGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTT 659
              | |||||||| |||||||| ||| |||||| || |||| ||||||||||||||||||||
Db      16789 GGCGCCAGCAGCGGCAACCCTCTCCGGTCCCTTTGCACATTGTGGAAGCTTTGTTCTTTC 16730

Qy        660 GCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCT 719
              ||| ||  ||| ||| |||||||||||||| |||||| ||||||||| ||||||||||||
Db      16729 GCTGTTTGCAACAAACCTTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCT 16670

Qy        720 GTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCA 778
              |||||| ||||  ||| ||||||  |||||| |||||||| |||| ||||||||||| |
Db      16669 GTAACAGTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCAACGAGACCACGAACCTA 16610

Qy        779 CTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAA 838
```

```
Db      16609 CAAGGAGGAACGAACGACTCCAGACGCGCCGCC-TTAAGAGCTGTAACACTCACCGCGAA 16551

Qy        839 GCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTC 898

Db      16550 GGTCCGCAGCTTCACTCTTG-AGCCAGCGAGACCACGAACCCACTAGAAAGAAGAAACTC 16492

Qy        899 TGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAA 958

Db      16491 CGAACACATCCGAACATCAGAAGGAACAAACTCCAGACACGCCGCCTTTAAGAACTGTAA 16432

Qy        959 CACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC-- 1016

Db      16431 CACTCACCGCGAGGGTCCGCGGCTTCATTTTTGAAGTCAGTGAGACCAAGAACCCACCAA 16372

Qy       1017 ----------------------GGAAGGAACAAATTCCAGACACAGTAGGAAATCTGTA 1053

Db      16371 TTCCGGACACACTCAGAGCCCAAGATAGCATCCTACCCAGCACAAGAATAGCTCACCTTG 16312

Qy       1054 TTTTTGATCTGTGGCTTCCAGGGTTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAG 1113

Db      16311 ATTGTCATTGCTGGTTGGTATTTAATGGGGAACCTTGACAGTTCCAACCCTTTTCCTTCA 16252

Qy       1114 GA---AACAGAGAATGGTTTGGAGGAGCACATGTGGGAATTGTTATGGACCAGGCTTGAG 1170

Db      16251 GGCCTATCAGTGAAAGCAGTTGTCCAGTTCATCTTGTAAGACCAGCAGGGTCTCCAGAGC 16192

Qy       1171 ATGCACATAGGGCA--------TTTCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAAT 1222

Db      16191 ATGGACAGACTACAGGGTAGGTTATCATTTGTCCTAAGGCTGGAAAATGTGGCATCACAG 16132

Qy       1223 ATAATCTAAGGAAGACAGTTTTTGTAGA--CAGTAGTAGTCTTTGCATCTGAGACATGTA 1280

Db      16131 GATTCATCCAAAAGGCAGCATTGGGAGAGTCTATACTAGTCAGCACTCCAAAGAGTGTAA 16072

Qy       1281 GATTATCAAGCAATTAATTAGAAAAAATATA----------------------GCCAGGTGC 1320

Db      16071 GTCAACCAGGAGAATCCAGGGGTCAGACCAAGGATCAAAAACCAGGGAGGGGCACAGCGT 16012

Qy       1321 GATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGG--GGTGTGGATCACGAGG 1378

Db      16011 GGTGGCTCATGCCTATAATCCCAGCCCTTTGGGAGACAGAGGCAGGAGGTTTTCTTGAGC 15952

Qy       1379 TCAGGCGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACA- 1437

Db      15951 CCAGCAGTTTGAGACCAGCCCGGGTAACATAGGGAGATCCTGTCTCTACAAAAAATAAAT 15892

Qy       1438 ------AAAATTAGCCTGGTGTGGTGGCACGCATCTGTAATCCCAG-TACTCAGGAGGCT 1490

Db      15891 AAAAATAAAAATAGCCAGGCATTGTGACACACTCCTATAGTCCCAGCTACTTGCAGAGCA 15832

Qy       1491 GAGGCAGGGGAATCTCTTGAACTTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCA 1550

Db      15831 GAGATGGGAGGATCATTTGAGCCCAGCAAGTAGAGGCTGCAGTGAGCTGTGATCATGCCA 15772

Qy       1551 CAGCACTCCATCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAGG 1610

Db      15771 CTGCACTCCAGCATGGGCAACAGAGGGAGACCTTGTCTCAAAAAAAAAAAAAAAAAATAC 15712

Qy       1611 AAAGGA 1616

Db      15711 CAAGGA 15706


RESULT 32
US-09-949-016-15347
; Sequence 15347, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15347
;   LENGTH: 80706
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(80706)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15347

  Query Match             18.5%;  Score 371;  DB 3;  Length 80706;
  Best Local Similarity   81.2%;  Pred. No. 3.1e-87;
  Matches  483;  Conservative    0;  Mismatches  100;  Indels   12;  Gaps    4;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490

Db      52296 TCTAGTGGGGAGGTGGAGAACTTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAGTCAG 52355

Qy        491 CACTCTGT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAA 542

Db      52356 CACCCTGTCAAAACAGACCAACCAGCTCTCTATAAAATGGACCAGTCGGCTCTCTGTAAA 52415

Qy        543 ATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGA 602

Db      52416 ATGGACCAATCAGCAGGATATGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCGAA 52475

Qy        603 GCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCT 662

Db      52476 TAACGCAGTGGCAACCCGCTCGGGTCCACTTCCGCACTGTGGAAACTTTGTTCTTTCTCT 52535

Qy        663 CTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTA 722

Db      52536 GTTTGCCATAAATCTTGCCACTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGTTGTA 52595

Qy        723 ACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTG 781

Db      52596 ACACTCACCGCGAAGGTCTGCAGCTTCACTCCTGAAGCCATCGAGACCACGAACCCACTG 52655

Qy        782 GAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCT 841

Db      52656 GGAGGAACAAACAACTCCAGAGGCTCCGCC-TTAAGAGCTGTAACACTCACCGCGAAGGC 52714

Qy        842 CTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGG 901

Db      52715 CTGCAGCTTCACTCCTGA--TCCAGGAGACCACAAACCCACCGGAAGGAAGAAACTCCTA 52772

Qy        902 ACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACAC 961

Db      52773 ACACATCCGAACATCAGAAGGAACAAACTCCGGACACACCGCCTTTAAGAACTATAACAC 52832

Qy        962 TCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016

Db      52833 TCACCTCGAGGGTCCGCGGCTTCATTCTTGCAGTCAGTGAGACCAAGAACCCACC 52887


RESULT 33
US-09-949-016-16105/c
; Sequence 16105, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16105
;   LENGTH: 52992
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16105

  Query Match             18.5%;  Score 370.4;  DB 3;  Length 52992;
  Best Local Similarity   78.3%;  Pred. No. 3.7e-87;
  Matches  515;  Conservative    0;  Mismatches   71;  Indels   72;  Gaps    3;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      45296 TCTAGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAGAGGATTGTAAATGCACCAATCAG 45237

Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGC---------------------------- 522
              |||||||||||||||| | ||||||||||||
Db      45236 CACTCTGTGTCTAGCTCAGGGATTGTAAATGCACCAGTCAGCACCCTGTCAAGACAGACC 45177

Qy        523 ----------------------ACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGAT 561
                                   |||||||||| |||||||||||| |||||| |||| |||
Db      45176 AAACAGCTCTCTGTAAAATGGACCAATCAGCTCTCTGTAAAATGCACCAATTAGCAAGAT 45117

Qy        562 GTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCAC 621
              |||||| |||| || |||||||| ||||| | |||  ||||||||||||||||||||||||
Db      45116 GTGGGTGGGGCCAGATAAGGGAATAAAAGCAGGCTGCCCGAGCCAGCAGTGGCAACCCAC 45057

Qy        622 TCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCT 681
              | |||||||||||||||||||| |||| ||||||||||  |||||  ||||||||||||||
Db      45056 TTGGGTCCCCTTCCACACTGTAGAAGTTTTGTTCTTTCACTCTTTGCAATAAATCTTGCT 44997

Qy        682 GCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCT 741
              |||||||| || ||| ||||||||| |||||||||||||||||||||| |||| |||||
Db      44996 GCTGCTCACTCCTTGGGTCCACACTGCCTTTATGAGCTGTAACACTCACCGCGAAGGTCT 44937

Qy        742 GTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCC 800
              |  |||||| |||||| | || | |||||||||| ||||  | |||||  || ||||||
Db      44936 GCAGCTTCACTCCTGAGGCCAGCGAGACCACGAATCCACCAGGAGGAATGAACAACTCCA 44877

Qy        801 GATGTGCTGCCT----------------------TTAAGAGCTGTAACACTCACTGCGAA 838
              |||| |  |  |                      |||||||||||||||||||| | |||
Db      44876 GATGGGAGGAATGAGCAACTCCAGACGCACCGCCTTAAGAGCTGTAACACTCACCGTGAA 44817

Qy        839 GCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTC 898
              | |||||||||||||||||||| ||| || |||||||||||| ||||| ||||||||||
Db      44816 GGTCTGCAGCTTCACTCCTGAAGCCAGCGAAACCACAAACCCATCAGAAAGAAGAAACTC 44757

Qy        899 TGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAA 958
              | |||   ||||| ||| |||||||||||||| |||||||||||||| ||| ||||||
Db      44756 CGAACATGTCTGAACATCAGAAGGAACAAACTCCGGACACACCATCTTTAAGAACTGTAA 44697

Qy        959 CACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016
              |||||||||||||||| | |||||||||||||||||||  ||||||||||||||||
Db      44696 CACTCACCGCAAGGGTCCGCGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCACC 44639


RESULT 34
US-09-949-016-15670/c
; Sequence 15670, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15670
;   LENGTH: 33578
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15670

  Query Match             18.4%;  Score 367.4;  DB 3;  Length 33578;
  Best Local Similarity   79.5%;  Pred. No. 1.9e-86;
  Matches  505;  Conservative    0;  Mismatches   81;  Indels   49;  Gaps    4;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||||||||||||||||||||| |||||||||| ||||||||||||||||||||||
Db      7467 TCTAGTGGGGACTTGGAGAACTTTTATGTCTAGCTAGAGGATTGTAAATGCACCAATCAG 7408

Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCA----------------- 533
              || |||||||||||||| | |||||||||| ||||||||||||
Db      7407 CATTCTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCACCCTGTCAAAACGGACC 7348

Qy        534 --------CTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGT 585
                       |||||||||||||||||||||||||||||| ||| |||| || |||||||| |
Db      7347 AATCAGCTCTCTGTAAAATGGACCAATCAGCAGGATGCGGGTGGGGCCAGATAAGGGAAT 7288

Qy        586 AAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGA 645
              |||| | |||  ||| |||||||||| |||||||  ||||||||||||  ||||||||||
Db      7287 AAAAGCAGGCTGCCCAAGCCAGCAGCGGCAACCTGCTCGGGTCCCCTTTGACACTGTGGA 7228

Qy        646 AGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCT-CATTCTTTGTGTCCACA 704
              ||||||||||||    ||||  |||||||||||||||||||| |   || || |||| ||
Db      7227 AGCTTTGTTCTTTCAGTCTTTGCAATAAATCTTGCTGCTGCTGCTCACTCTGGGTCCGCA 7168

Qy        705 CTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC 764
              || || |||||||||||||||||||||||| ||||||  ||||||  |||||| | || |
Db      7167 CTGCCGTTATGAGCTGTAACACTCACTGCGAAGGTCTGCAGCTTCACTCCTGAGGCCAGC 7108

Qy        765 -AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGC------------- 810
               |||||||||||||| || |||||  || |||||| ||    | |
Db      7107 GAGACCACGAACCCACCGGGAGGAATGAACAACTCCAGACAGGAGGAACTAACAACTCCA 7048

Qy        811 ---------CTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAG 861
                        | |||||||||||||||||||||  |||||  |||||||||||||||||||
Db      7047 GATGCGCCGCCTTAAGAGCTGTAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAG 6988

Qy        862 TCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAG 921
              |||  ||||||| |||||||||||||||||||||| | |||   | ||| ||| ||||
Db      6987 CCAGCAAGACCACGAACCCACCAGAAGGAAGAAACTCCGAACATGTCCGAACATCAGAAG 6928

Qy        922 GAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGC 981
              |||||||||| |||||||||||||| ||| |||||||||||| ||||||||| | |||
Db      6927 AAACAAACTCCGGACACACCATCTTTAAGAACTGTAACACTCACTGCAAGGGTCCGCGGC 6868

Qy        982 TTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016
              |||||||||||||||  ||||| ||||||||||
Db      6867 TTCATTCTTGAAGTCAGTGAGACCGAGAACCCACC 6833


RESULT 35
US-09-949-016-16041/c
; Sequence 16041, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
```

```
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16041
;   LENGTH: 87644
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16041

  Query Match             18.3%;  Score 365.4;  DB 3;  Length 87644;
  Best Local Similarity   82.6%;  Pred. No. 9.8e-86;
  Matches  493;  Conservative    0;  Mismatches   91;  Indels   13;  Gaps    6;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||| ||| |||||||||||||||||||||| | |||||||||| ||||||||||
Db      34681 TCTAGTGGTGACGTGGAGAACTTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAG 34622

Qy        491 CACTCTGT--------GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAA 542
              |||  |||          | | |       |   |||   ||||||| || ||||||||||
Db      34621 CACCGTGTCAAAATGGACCAATCAGCGCTCTGTAAAACAGACCAATCGGCTCTCTGTAAA 34562

Qy        543 ATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGA 602
              ||||  |||||||||||||||||||| |||| || |||||  | |||||  | |||  |||||
Db      34561 ATGGGCCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAAAATAAAAGCAGGCTGCCCGA 34502

Qy        603 GCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCT 662
              |||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||||
Db      34501 GCCAGCAGTGGCAACCCGCTCGGGTCCCCTTCCATACTGTGGAAGCTTTGTTCTTTTGCT 34442

Qy        663 CTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTA 722
              ||||  |||||||||||||||||||||||| |||||| |||||||| ||||||||||||||||
Db      34441 CTTCGCAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCACACTGCCTTTATGAGCTGTA 34382

Qy        723 ACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTG 781
              ||||||||  |||| ||||||  |||||| |||||||| || | ||||||||||||||| |
Db      34381 ACACTCACCGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACCACGAACCCACCG 34322

Qy        782 GAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCT 841
              | ||||| |||  |||||  ||    |||||  |||||||||||||||||||  |||||| |
Db      34321 GGAGGAAAAAAACGACTCCAGACTCACTGCC-TTAAGAGCTGTAACACTCATCGCGAAGGT 34263

Qy        842 CTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGG 901
              ||||||||||  |||||  ||  ||||  ||||||  |||||||||  ||||||||||  |
Db      34262 CTGCAGCTTCAATCCTG-AGCCAGTAAGACCATAAACCCACCAAAAGGAAGAAACTTCGA 34204

Qy        902 ACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTG-TAACA 960
              |||||  |   | |||  ||||||||||||||  ||   ||  ||||  |||  |||||
Db      34203 ACACATCCCCACATCAGAAGGAACAAACTCTGGACATGCCGCCTTTAAGAACTGTTAACA 34144

Qy        961 CTCACCGCAAGGGTCTGTGGCTTCATTCTT-GAAGTCAGCAAGACCAAGAACCCACC 1016
              |||||||||||||||  ||||||||||||||  |||||||    ||||||||||||||||
Db      34143 CTCACCGCAAGGGTCCGTGGCTTCATTCTTGGAAGTCATTGAGACCAAGAACCCACC 34087


RESULT 36
US-09-422-576D-6
; Sequence 6, Application US/09422576D
; Patent No. 6395549
; GENERAL INFORMATION:
;  APPLICANT: Tuan, Dorothy
;  APPLICANT:  Long, Qiaoming
;  APPLICANT:  Bengra, Chikh
;  TITLE OF INVENTION: Long Terminal Repeat, Enhancer, and Insulator Sequences for Use i
;  TITLE OF INVENTION:  Recombinant Vectors
;  FILE REFERENCE: M0351-205010
;  CURRENT APPLICATION NUMBER: US/09/422,576D
;  CURRENT FILING DATE:  1999-10-21
;  PRIOR APPLICATION NUMBER: US 60/105,256
;  PRIOR FILING DATE: 1998-10-22
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 1043
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-422-576D-6

  Query Match             18.1%;  Score 361.8;  DB 3;  Length 1043;
  Best Local Similarity   80.5%;  Pred. No. 1.2e-85;
  Matches  466;  Conservative    0;  Mismatches   97;  Indels   16;  Gaps    3;

Qy        453 TTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGA 512
              | ||| ||||||| |||| ||||||||  |||||||||||||| |||||||||||| | |||
Db        359 TCTGTATCTAGCTCAAGGTTTGTAAACACACCAATCAGCACCCTGTGTCTAGCTCATGGA 418

Qy        513 TTGTAAATGCACCAATCAGCACTCTGTAAAA--------------TGGACCAATCAGCAG 558
              ||||||||||||||||||  | |||| |||                |  ||||||||||||
Db        419 TTGTAAATGCACCAATCAGTGCCCTGTCAAAACAGACCACTGGGCTCTACCAATCAGCAG 478

Qy        559 GATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACC 618
              |||||||| |||| || ||||| || ||||| | |||  ||||||||||||||||||||||
Db        479 GATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCGAGCCAGCAGTGGCAACC 538

Qy        619 CACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTT 678
              | |||||||||||||||||||||||||||||||||||||||||| ||||||  ||||||||||
Db        539 CGCTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTCGCTCTTTGCAATAAATCTT 598

Qy        679 GCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGG 738
              |||||||||| | |||| ||| ||||| |||||||||||||||||| ||||| |||| ||
Db        599 GCTGCTGCTCACTGTTTGGGTCTACACTGCCTTTATGAGCTGTAACGCTCACCGCGAAGG 658

Qy        739 TCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACT 797
              ||||  |||||| || ||||| || | |||||||||||||| || |||||| || ||||
Db        659 TCTGCAGCTTCACTCTTGAAGCCAGCGAGACCACGAACCCACCGGGAGGAACGAACAACT 718

Qy        798 CCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCT 857
              || || | || ||||| |       | |          |  ||| ||||||||||||||||
Db        719 CCAGAGGCGCCGCCTTAAGAGCTGGAACGTTCACTGGTAAAGGTCTGCAGCTTCACTCCT 778

Qy        858 GAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCT 917
              | || ||| |||||||| |||||| |||||||||||||||| | ||||| |  || |||
Db        779 G-AGCCAGCGAGACCACGAACCCATCAGAAGGAAGAAACTCCGAACACATCCAAACATCA 837

Qy        918 GAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTG 977
              ||| |||||||||||| |||| |    |||| ||| |||||||||||||| | ||||||
Db        838 GAACGAACAAACTCCACACACGCAGCCTTTAAGAACTGTAACACTCACCACGAGGGTCCC 897

Qy        978 TGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016
              ||||||||||||||||||  | ||||||||||||||
Db        898 CGGCTTCATTCTTGAAGTCAGTGAAACCAAGAACCCACC 936


RESULT 37
US-09-949-016-11817/c
; Sequence 11817, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
```

```
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11817
;   LENGTH: 124700
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-11817

  Query Match             18.0%;  Score 359.8;  DB 3;  Length 124700;
  Best Local Similarity   80.4%;  Pred. No. 3.4e-84;
  Matches  469;  Conservative    0;  Mismatches   87;  Indels   27;  Gaps    3;

Qy        464 CTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCA 523
              ||  |  || |   |  || |  || | |     |  | | |   |   |  |||| |
Db      30935 CTGTAAAATGGACCAATCAGCTCTCTGTAAAAAGGACCAATCAGCACTCTGTAAAATGGA 30876

Qy        524 CCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGA 583
              ||||||||| ||||| |||||||||||||||||||||||||||||| ||| ||||||||||
Db      30875 CCAATCAGCTCTCTGCAAAATGGACCAATCAGCAGGATGTGGGCAGGGCCAAATAAGGGA 30816

Qy        584 GTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTG 643
               |||||  |||||||| |||||||| |||||||| ||| | ||||||||||||||||
Db      30815 ATAAAAGCTGGCCACCTGAGCCAGCAGCGGCAACCCGCTCAGATCCCCTTCCACACTGTG 30756

Qy        644 GAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCAC 703
              ||| ||||||||||||||||||||||||||||||||||||||||||||| |||||| |||  |
Db      30755 GAATCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCTGC 30696

Qy        704 ACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA- 762
              |||||||||||||||| |||||||||| ||||||||| ||||||| |||||| ||||
Db      30695 ACTACCTTTATGAGCTGAAACACTCACTGAGAGGGTCTGCGGCTTCACTCCTGAGGTCAG 30636

Qy        763 ACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTG 822
              ||||||||||||||| |||||||||||  |||||  |||| ||  ||||||||||||
Db      30635 TGAGACCACGAACCCACCGGAAGGAACAAACAACTCTGGATGCGCCACCTTTAAGAGCTG 30576

Qy        823 TAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCAC 882
              |||||||||||| |||| | ||| |||||||||||||||||| ||||||||||||||||
Db      30575 TAACACTCACTGGGAAGGTTTGCGGCTTCACTCCTGAAGTCAGCGAGACCACAAACCCAC 30516

Qy        883 CAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCA 942
              |||||||||||||||  |||||        ||||||||||||||||||  |||||||||
Db      30515 TGGAAGGAAGAAACTCCAGACAC--------ATCTGAAGGAACAAACTCTGGACACACCA 30464

Qy        943 TCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAG 1002
              ||||| ||| |||||||||||||||||||||                    |||  ||
Db      30463 TCTTTAAGAACTGTAACACTCACCGCAAGGGTC------------------CACAGTGAG 30422

Qy       1003 ACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
              |||||||||||||||||| ||| ||||||| ||||||  || |
Db      30421 ACCAAGAACCCACCGGAATGAATAAATTCCGGACACAAAAGAA 30379


RESULT 38
US-09-949-016-15439/c
; Sequence 15439, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15439
;   LENGTH: 124701
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15439

  Query Match             18.0%;  Score 359.8;  DB 3;  Length 124701;
  Best Local Similarity   80.4%;  Pred. No. 3.4e-84;
  Matches  469;  Conservative    0;  Mismatches   87;  Indels   27;  Gaps    3;

Qy        464 CTAAAGGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCA 523
              ||  |  || |   |  || |  || | |     |  | | |   |   |  |||| |
Db      30935 CTGTAAAATGGACCAATCAGCTCTCTGTAAAAAGGACCAATCAGCACTCTGTAAAATGGA 30876

Qy        524 CCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGA 583
              ||||||||| ||||| |||||||||||||||||||||||||||||| ||| ||||||||||
Db      30875 CCAATCAGCTCTCTGCAAAATGGACCAATCAGCAGGATGTGGGCAGGGCCAAATAAGGGA 30816

Qy        584 GTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTG 643
               |||||  |||||||| |||||||| |||||||| ||| | ||||||||||||||||
Db      30815 ATAAAAGCTGGCCACCTGAGCCAGCAGCGGCAACCCGCTCAGATCCCCTTCCACACTGTG 30756

Qy        644 GAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCAC 703
              ||| ||||||||||||||||||||||||||||||||||||||||||||| |||||| |||  |
Db      30755 GAATCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCACTCTTTGGGTCTGC 30696

Qy        704 ACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA- 762
              |||||||||||||||| |||||||||| ||||||||| ||||||| |||||| ||||
Db      30695 ACTACCTTTATGAGCTGAAACACTCACTGAGAGGGTCTGCGGCTTCACTCCTGAGGTCAG 30636

Qy        763 ACAGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTG 822
              ||||||||||||||| |||||||||||  |||||  |||| ||  ||||||||||||
Db      30635 TGAGACCACGAACCCACCGGAAGGAACAAACAACTCTGGATGCGCCACCTTTAAGAGCTG 30576

Qy        823 TAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCAC 882
              |||||||||||| |||| | ||| |||||||||||||||||| ||||||||||||||||
Db      30575 TAACACTCACTGGGAAGGTTTGCGGCTTCACTCCTGAAGTCAGCGAGACCACAAACCCAC 30516

Qy        883 CAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCA 942
              |||||||||||||||  |||||        ||||||||||||||||||  |||||||||
Db      30515 TGGAAGGAAGAAACTCCAGACAC--------ATCTGAAGGAACAAACTCTGGACACACCA 30464

Qy        943 TCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAG 1002
              ||||| ||| |||||||||||||||||||||                    |||  ||
Db      30463 TCTTTAAGAACTGTAACACTCACCGCAAGGGTC------------------CACAGTGAG 30422

Qy       1003 ACCAAGAACCCACCGGAAGGAACAAATTCCAGACACAGTAGGA 1045
              |||||||||||||||||| ||| ||||||| ||||||  || |
Db      30421 ACCAAGAACCCACCGGAATGAATAAATTCCGGACACAAAAGAA 30379


RESULT 39
US-09-949-016-12372/c
; Sequence 12372, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12372
```

```
;   LENGTH: 181429
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12372

  Query Match             17.9%;  Score 357.2;  DB 3;  Length 181429;
  Best Local Similarity   81.8%;  Pred. No. 2e-83;
  Matches  481;  Conservative    0;  Mismatches   83;  Indels   24;  Gaps    5;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||||||| |||||||||||||||||||||| | ||||||||||  ||||||||||
Db     138282 TCTAGTGGGGACATGGAGAACTTTTGTGTCTAGCTCAGGGATTGTAAACACACCAATCAG 138223

Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCA 550
              ||| ||||                  |||   |||||||||| ||||||||||  |||||
Db     138222 CACCCTGTCA----------------AAACAGACCAATCAGCTCTCTGTAAAACAGACCA 138179

Qy        551 ATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAG 610
              |||||||||||||||| |||| || ||||| || ||||| | |||  ||||||||||||
Db     138178 ATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCGAGCCAGCAG 138119

Qy        611 TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAA 670
              |||||| || ||| |||||||||||||||||||||||||     |||  |||||     |
Db     138118 TGGCAATCCTCTCAGGTCCCCTTCCACACTGTGGAAGCTT-----TTTCACTCTTTGTGA 138064

Qy        671 TAAATCTTGCTGCTGCTCATTC-TTTGTGTCCACACTACCTTTATGAGCTGTAACACTCA 729
              |||||||||||||||||| || |||| ||||||| | ||||||||||||||||||| |||
Db     138063 TAAATCTTGCTGCTGCTCACTCTTTTGGGTCCACATTGCCTTTATGAGCTGTAACAATCA 138004

Qy        730 CTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAA 788
              |||||  ||||||  |||||| |||||||| || | ||||||||||||||  |||||||
Db     138003 ATGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCAAGACCACGAACCCACCAGAAGGAA 137944

Qy        789 CAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGC 848
              | ||  ||||||  || |   ||||| |||||||||||||||||||| ||||||  |||||||
Db     137943 CGAACAACTCCAGACGCACTGCC-TTAAGAGCTGTAACACTCACCGCGAAGGTCTGCAGC 137885

Qy        849 TTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACC 908
              |||||||  |||| ||| |||||||   | ||||||||||||||||||||   ||||| |
Db     137884 TTCACTCCCGAAGCCAGCGAGACCAGGAGCCCACCAGAAGGAAGAAACTCCAAACACATC 137825

Qy        909 TGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGC 968
              |||||||  |||| |||||||||  ||||| ||| |||| | | ||||||||||||| |
Db     137824 CGAATATCAGAAGAAACAAACTCTGGACACGCCACCTTTAAAAACTGTAACACTCACCAC 137765

Qy        969 AAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016
              ||||||  | ||||||||||||||||||  |||||||||||| ||
Db     137764 GAGGGTCCGCGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCCCC 137717


RESULT 40
US-09-949-016-15772/c
; Sequence 15772, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15772
;   LENGTH: 181430
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15772

  Query Match             17.9%;  Score 357.2;  DB 3;  Length 181430;
  Best Local Similarity   81.8%;  Pred. No. 2e-83;
  Matches  481;  Conservative    0;  Mismatches   83;  Indels   24;  Gaps    5;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||||||| |||||||||||||||||||||| | ||||||||||  ||||||||||
Db     138282 TCTAGTGGGGACATGGAGAACTTTTGTGTCTAGCTCAGGGATTGTAAACACACCAATCAG 138223

Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGGACCA 550
              ||| ||||                  |||   |||||||||| ||||||||||  |||||
Db     138222 CACCCTGTCA----------------AAACAGACCAATCAGCTCTCTGTAAAACAGACCA 138179

Qy        551 ATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAG 610
              |||||||||||||||| |||| || ||||| || ||||| | |||  ||||||||||||
Db     138178 ATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGGCTGCCCGAGCCAGCAG 138119

Qy        611 TGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAA 670
              |||||| || ||| |||||||||||||||||||||||||     |||  |||||     |
Db     138118 TGGCAATCCTCTCAGGTCCCCTTCCACACTGTGGAAGCTT-----TTTCACTCTTTGTGA 138064

Qy        671 TAAATCTTGCTGCTGCTCATTC-TTTGTGTCCACACTACCTTTATGAGCTGTAACACTCA 729
              |||||||||||||||||| || |||| ||||||| | ||||||||||||||||||| |||
Db     138063 TAAATCTTGCTGCTGCTCACTCTTTTGGGTCCACATTGCCTTTATGAGCTGTAACAATCA 138004

Qy        730 CTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAAGGAA 788
              |||||  ||||||  |||||| |||||||| || | ||||||||||||||  |||||||
Db     138003 ATGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCAAGACCACGAACCCACCAGAAGGAA 137944

Qy        789 CAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGC 848
              | ||  ||||||  || |   ||||| |||||||||||||||||||| ||||||  |||||||
Db     137943 CGAACAACTCCAGACGCACTGCC-TTAAGAGCTGTAACACTCACCGCGAAGGTCTGCAGC 137885

Qy        849 TTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGAAGAAACTCTGGACACACC 908
              |||||||  |||| ||| |||||||   | ||||||||||||||||||||   ||||| |
Db     137884 TTCACTCCCGAAGCCAGCGAGACCAGGAGCCCACCAGAAGGAAGAAACTCCAAACACATC 137825

Qy        909 TGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAGAGCTGTAACACTCACCGC 968
              |||||||  |||| |||||||||  ||||| ||| |||| | | ||||||||||||| |
Db     137824 CGAATATCAGAAGAAACAAACTCTGGACACGCCACCTTTAAAAACTGTAACACTCACCAC 137765

Qy        969 AAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACC 1016
              ||||||  | ||||||||||||||||||  |||||||||||| ||
Db     137764 GAGGGTCCGCGGCTTCATTCTTGAAGTCAGTGAGACCAAGAACCCCCC 137717


RESULT 41
US-09-949-016-13203
; Sequence 13203, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13203
;   LENGTH: 94019
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(94019)
```

```
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13203

  Query Match             17.7%;  Score 355;  DB 3;  Length 94019;
  Best Local Similarity   60.4%;  Pred. No. 5.6e-83;
  Matches  889;  Conservative    0;  Mismatches  425;  Indels  159;  Gaps   12;

Qy        409 TGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAA 468
              || | | |||||  ||  |||| |    | | || ||           | | || ||||
Db      88223 TGTATCTAGCTGCTCTGGTGGGGCCTTGGAGAACCTGTGTGTGGAAACTCTGTATCTAAC 88282

Qy        469 GGATTGTAAATGCACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAAT 528
                ||   |  | ||       || ||| |||||| | |||||||||||| ||||||
Db      88283 TAATCTGATGGGGACGTGGAGAACCTTTGTATCTAGCTCAGGGATTGTAAATGTACCAAT 88342

Qy        529 CAGCACTCTGTAAAA--------------TGGACCAATCAGCAGGATGTGGGCGGGGTCA 574
              |||| | |||  |||               |||||||||||||||||||| |||| ||
Db      88343 CAGCGCCCTGACAAAACAGGCCACTGGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCA 88402

Qy        575 AATAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTC 634
              ||||| || ||||| | |||  |||||||| ||| ||||||||||||||||||||||||
Db      88403 GATAAGAGAATAAAAGCAGGCTGCCCGAGCCCGCATTGGCAACCCACTCGGGTCCCCTTC 88462

Qy        635 CACACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTT 694
              |||||||||||| |||||||||||| ||||||  |||||| |||||||  |||||| ||||
Db      88463 CACACTGTGGAGGCTTTGTTCTTTCGCTCTTTGCAATAACTCTTGCTACTGCTCACTCTT 88522

Qy        695 TGTGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCC 754
              || |||||| || | |||||||| |||||||||||||||| | ||||  |||||| |||
Db      88523 TGGGTCCACGCTGCTTTTATGAGTTGTAACACTCACTGCGAAGATCTGCAGCTTCACTCC 88582

Qy        755 TGAAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTT 813
              ||||  || | ||||||||| |||||  | ||||| ||| |||||| || |||||||| |
Db      88583 TGAACCCAGCGAGACCACGAGCCCACCAGGAGGAATAAACAACTCCAGACGTGCTGCC-T 88641

Qy        814 TAAGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCA 873
              |||||||||||||||||| || ||| |||||||||||||||| || ||| |||||||
Db      88642 TAAGAGCTGTAACACTCACCGCAAAGGTCTGCAGCTTCACTCCTG-AGCCAGCGAGACCA 88700

Qy        874 CAAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCA 933
              ||||||||||||||||||||| | ||||| ||||| ||| ||||| ||| ||||||
Db      88701 TGAACCCACCAGAAGGAAGAAACTCCGAACACATCTGAACATCAGAAGGGACAGACTCCA 88760

Qy        934 GACACACCATCTTTCAGAGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
              ||     ||| | || ||||||||||     |||||| |||||| | ||||||||||||||
Db      88761 GATGTGCCA-CCTTAAGAGCTGTAA----CACCGCGAGGGTCCGCGGCTTCATTCTTGAA 88815

Qy        994 GTCAGCAAGACCAAGAACCCACC----------------------------------G 1017
              |||||  |||||||| |||||||
Db      88816 GTCAGTGAGACCAAGGACCCACCAATTCCGGACACACCAATACGTGGATAGGGATAGATC 88875

Qy       1018 GAAGGAACAAATTCCAGACACAGTAGGAAATCTGTATTTTTGATCTGTGGCTTCCAGGG- 1076
              |||| |    |  | | | | || | |      |  |      |    |  |||||  |
Db      88876 AAAGGGAATCAAACTAAAGATAGGAAGGCCAGTAAAGAGGCAACTGCAGTAATCCAGAGG 88935

Qy       1077 ----------------TTACTCCAGTCATTGAAGTCTCCATTGCAGCCTTAAGGAAACAG 1120
                           | | |  ||| | ||     | || |  ||     ||| | ||
Db      88936 AAGAGAGATGGAAGCTTAAATTAAGTAAATGCCAGGTGCAATAAAGAGGAGAGGGAGCAC 88995

Qy       1121 AGAATGGTTTGGAGGAG-------CACATGTGGGAATTGTTATGGACCAGGCTTGAGATG 1173
                ||     ||||||       |||| |   |  |   |      |||   | ||
Db      88996 CAGGTGAGAGTGAGGAGGAGAAGCCACAGGCCTGGGTGCCTGAATCTCAGTTGTCAGGCC 89055

Qy       1174 CACATAGGGCATTTCTGATCAAACCTAGCTGGAAGCAGGGCCAGGAAATATAATCTAAGG 1233
                ||   | |  ||    | |||   |    || | ||  | | |   ||     ||
Db      89056 TGGGCAGTTGAGTGGTGTGGATACCACTCACCGAGAATGGGAACGCCAGGAAAGAAGTGG 89115

Qy       1234 AAGACAGTTTTTGTAGACAGTAGTAGTCTTTGCATCTGAGACATGTAGATTATCAAGCAA 1293
              |          || |      ||||  |    | |      || | |   ||   |
Db      89116 ATTTGGAGAAACGTGGGGGAGCTTAGTTCTGTGGTGTCTTTGATTTTGGACATACTGGGT 89175

Qy       1294 TTAATTAGAAAAAATA---------------------------------------------- 1309
              || |  |||| |   |
```

```
Db      89176 TTGAAGAGAACAGCCAATTACTGATGCTTAGTAGGTAATGGCTTGGTGCTTCAGGAAGAG 89235

Qy       1310 ----------------------------TAGCCAGGTGCGATGGCTCATGCCTGTAATC 1340
                                          | ||  || ||| ||||||| ||||||||||
Db      89236 ATCAGAACCCTTTGATAAGAACCCTGGCCTGGCTGGGCGCGGTGGCTCACGCCTGTAATC 89295

Qy       1341 CCAGCACTTTGGGAGGCCAAGGGGTGTGGATCACGAGGTCAGGCGTTCGAGACCAGCCTG 1400
              |||||||||| |||||| ||| | | ||||||||||||||||| | || ||||||  ||||
Db      89296 CCAGCACTTTGAGAGGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCAAGACCATCCTG 89355

Qy       1401 GCCAACATGGTGAAACCCCGTCTCTACTAAAAATA---CAAAAATTAGCCTGGTGTGGTG 1457
              || |||| |||||||||| ||||||||||||||       |||||||||||| || | ||||
Db      89356 GCTAACACGGTGAAACCCAGTCTCTACTAAAAAATACCAAAAAATTAGCCGGGCGCGGTG 89415

Qy       1458 GCACGCATCTGTAATCCCAG-TACTCAGGAGGCTGAGGCAGGGGAATCTCTTGAACTTGG 1516
              ||  |   ||||| |||||| |||  ||| |||||||||| ||||  |  ||||  ||
Db      89416 GCGGGAGCCTGTAGTCCCAGCTACTGGGGAAGCTGAGGCAGGAGAATGGCGGGAACCCGG 89475

Qy       1517 GAGGCAGAGGTTGCAGTGAGCCAAGATCACACCACAGCACTCCATCCTGGGTGACAGAGC 1576
              ||||| ||| |||| || |||| ||||||| |||| |||||||| ||||||||||||||
Db      89476 GAGGCGGAGCTTGCTGTAAGCCGAGATCACGCCACTGCACTCCAGCCTGGGTGACAGAGC 89535

Qy       1577 GAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGGAAAGGAAAATATAATCAAGAATATTG 1636
              |||||||  |||||||||||| |||| ||| |   ||  |||  |   |     || | |
Db      89536 GAGACTCCATCTCAAAAAAAATAAAATAAATAAATAAATAAAGAACCCTGGCCCATTTAG 89595

Qy       1637 ACAGGTAACATTTATTCAACACTTACTATGCACCAGGCAATACACTAAGTGTTTTACATG 1696
              | | | | |  |    |     | |||   | ||  || |          | ||  |
Db      89596 AAATGAAGCTGTCTATACTTCCAGACTGAATAACAACCACTTTTTGTTAGTTGTTTAAGA 89655

Qy       1697 GATTAACTCATTTAATCTTAACAATAGCCCTAT 1729
                | | |||          ||| | |||| | |||
Db      89656 TGTGATCTCTGCCTTAATTATCGATAGTCATAT 89688
```

```
RESULT 42
US-09-949-016-17028/c
; Sequence 17028, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17028
;   LENGTH: 88557
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(88557)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17028

  Query Match             17.5%;  Score 350.2;  DB 3;  Length 88557;
  Best Local Similarity   79.3%;  Pred. No. 1e-81;
  Matches  488;  Conservative    0;  Mismatches   98;  Indels   29;  Gaps    5;

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||  |||||||||||||||||||||||||||||||| |  |||||||| ||||||||||
Db      78270 TCTAGTGGGGACTTGGAGAACTTTTGTGTCTAGCTCAGAGATTGTAAACGCACCAATCAG 78211

Qy        491 CACTCTGT----GTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAAATGG 546
```

```
Db      78210 CACCCTGTCAAAGCAGACCAATCAGCTCTCTGTAAAACCAATCAGCTCTCTGTAAAATGC 78151

Qy        547 ACCAATCAGCAGGATGTGGGCGGGGTCAAATAAGGGAGTAAAAA-CTGGCCACCCGAGCC 605
Db      78150 ACCAATCAGTAACATGTGGGTGGGGCCAGATAAAGGAATAAAAAGCAGGCTGCCAGAGCC 78091

Qy        606 AGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGCTCTT 665
Db      78090 AGCAGTGGCAACCTGCTCAGGCCCCCTTTGACACTGTGGAAGCTTTGTTCTTTCACTCCT 78031

Qy        666 CACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTTTATGAGCTGTAACA 725
Db      78030 TGCAATAAATCTTGCTGCTGCTCACTCTTTGGGTCCGCACTGCCTTTGTGAGCTGTAACA 77971

Qy        726 CTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCAAC-AGACCACGAACCCACTGGAA 784
Db      77970 CTCACCGCGAAGGTCTGCAGCTTCACTCCTGAGGCCAGCAAAACCACGAACCCACCAGAA 77911

Qy        785 GGAACAAAGAACTCCCGATGTGCTGC----------------------CTTTAAGAGCTG 822
Db      77910 GGAATAAACAACTCCGGACGGGAGGAACGAACAGCTCCAGACGTGCTGCCTTAAAAGCTG 77851

Qy        823 TAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCAC 882
Db      77850 TAACACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGAGACCACGAACCCAC 77791

Qy        883 CAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCA 942
Db      77790 CAGAAGGAAGAAACTCTGAACACGTCCGAACATCAGAAGGAACAAACTCCGGATACACCA 77731

Qy        943 TCTTTCAGAGCTG-TAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAA 1001
Db      77730 TCTTTAAGAACTGTTAACACTCACCGTGAGGGTCCGCGGCTTCATTCTTGAAGTCAGTGA 77671

Qy       1002 GACCAAGAACCCACC 1016
Db      77670 GACCAAGAACCCACC 77656
```

RESULT 43
US-09-949-016-16179
; Sequence 16179, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16179
; LENGTH: 157644
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(157644)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16179

Query Match 17.5%; Score 349.2; DB 3; Length 157644;
Best Local Similarity 82.7%; Pred. No. 2.4e-81;
Matches 435; Conservative 0; Mismatches 88; Indels 3; Gaps 3;

```
Qy        517 AAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAA 576
Db      31518 AAATTGACCAATCAGTTCTCTGTAAAATGGACCAATCAGCAGGATCAGGGTGGGGCCAGA 31577

Qy        577 TAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA 636
Db      31578 TAAGGGAATAAACGCAGGCAACCTGAGCCAGCAGCAGCAACTCACTCGGGTACCCTTGCG 31637

Qy        637 CACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTG 696
Db      31638 TGTTGTGGAAGCTTTGTTCTTTCATTCTTCGCAATAAATCTTGCTGCTGCTCACTCTTTG 31697

Qy        697 TGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTG 756
Db      31698 GGTTCGCTCTGCCTTTATGAGCTGTAACACTCACTGCGAAGTTCTGCAGCTTCACTCCTG 31757

Qy        757 AAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTA 815
Db      31758 AAGCCAACAAGACCATGAACCCTCTGGGCAGAAGGAACAACTCTGGACACGCCACCTTGA 31817

Qy        816 AGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAG-TGAGACCAC 874
Db      31818 AGAGCTGTGACACTGACCGCGAAGGTCTGCAGCTTCACTCCTGAAATCAGCATTGACCAC 31877

Qy        875 AAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAG 934
Db      31878 GAACCCACCAGAAGGAAGATACTCCGGACACATCTGAACAGCTGAAGGAACAAACTCTGG 31937

Qy        935 ACACACCATCTTTCAGAGCTG-TAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
Db      31938 ACACACCATCTTTAAGAACTGTTAACACTCACCGCGAGGGTCCGCAGCTTCATTCTTGAA 31997

Qy        994 GTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACA 1039
Db      31998 GTCCGCGAGACCAAGAACCCACTCTAAGTAATCAATTCTGGACACA 32043
```

RESULT 44
US-09-949-016-16180
; Sequence 16180, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16180
; LENGTH: 157644
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(157644)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16180

Query Match 17.5%; Score 349.2; DB 3; Length 157644;
Best Local Similarity 82.7%; Pred. No. 2.4e-81;
Matches 435; Conservative 0; Mismatches 88; Indels 3; Gaps 3;

```
Qy        517 AAATGCACCAATCAGCACTCTGTAAAATGGACCAATCAGCAGGATGTGGGCGGGGTCAAA 576
Db      31518 AAATTGACCAATCAGTTCTCTGTAAAATGGACCAATCAGCAGGATCAGGGTGGGGCCAGA 31577
```

```
Qy        577 TAAGGGAGTAAAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCA 636
              ||||||| ||||  | ||| ||| ||||||||||  ||||| ||||||||| ||||| |
Db      31578 TAAGGGAATAAACGCAGGCAACCTGAGCCAGCAGCAGCAACTCACTCGGGTACCCTTGCG 31637

Qy        637 CACTGTGGAAGCTTTGTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTG 696
              ||||||||||||||||||    ||||| |||||||||||||||||||||| ||||||
Db      31638 TGTTGTGGAAGCTTTGTTCTTTCATTCTTCGCAATAAATCTTGCTGCTGCTCACTCTTTG 31697

Qy        697 TGTCCACACTACCTTTATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTG 756
              || | || |||||||||||||||||||||||||||| | ||||  |||||| |||||
Db      31698 GGTTCGCTCTGCCTTTATGAGCTGTAACACTCACTGCGAAGTTCTGCAGCTTCACTCCTG 31757

Qy        757 AAGTCAAC-AGACCACGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTA 815
              ||| |||| |||||| |||||| ||||    |||  || |||||  ||    ||  |||| |
Db      31758 AAGCCAACAAGACCATGAACCCTCTGGGCAGAAGGAACAACTCTGGACACGCCACCTTGA 31817

Qy        816 AGAGCTGTAACACTCACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAG-TGAGACCAC 874
              ||||||| ||||| || |||||| |||||||||||||||||||| ||||    ||||||
Db      31818 AGAGCTGTGACACTGACCGCGAAGGTCTGCAGCTTCACTCCTGAAATCAGCATTGACCAC 31877

Qy        875 AAACCCACCAGAAGGAAGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAG 934
              |||||||||||||||||| |||| ||||||| ||||| | ||||||||||||||||  |
Db      31878 GAACCCACCAGAAGGAAGATACTCCGGACACATCTGAACAGCTGAAGGAACAAACTCTGG 31937

Qy        935 ACACACCATCTTTCAGAGCTG-TAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAA 993
              |||||||||||||| ||| ||| |||||||||||||| |||||| |  ||||||||||||||
Db      31938 ACACACCATCTTTAAGAACTGTTAACACTCACCGCGAGGGTCCGCAGCTTCATTCTTGAA 31997

Qy        994 GTCAGCAAGACCAAGAACCCACCGGAAGGAACAAATTCCAGACACA 1039
              ||| || ||||||||||||||||   ||| ||  |||||  ||||||
Db      31998 GTCCGCGAGACCAAGAACCCACTCTAAGTAATCAATTCTGGACACA 32043
```

RESULT 45
US-09-949-016-14659/c
; Sequence 14659, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14659
; LENGTH: 22888
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14659

Query Match 17.4%; Score 348.6; DB 3; Length 22888;
Best Local Similarity 73.4%; Pred. No. 1.5e-81;
Matches 550; Conservative 0; Mismatches 119; Indels 80; Gaps 5;

```
Qy        371 GCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGG 430
              ||||      |||| ||  |   |      | ||||||||||||  ||||||||||||
Db      15993 GCAGGACCACTGTCTTGGCTCCTGATCAAAATGAGAGGTGAAGCTGGCTGGGCTTCTGGG 15934

Qy        431 TCAGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAG 490
              ||||||| |||||||||||||||| || |||||||||||||||||||||| ||||||
Db      15933 TCAGGTGAGGACTTGGAGAACTTTTCTGCCTAGCTAAAGGATTGTAAATGCACAAATCAG 15874

Qy        491 CACTCTGTGTCTAGCTAAAGGATTGTAAATGCACCAATCAGCACTCTGTAAA-------- 542
              | |||||||||||||||||| ||||||| ||||||||||||||||||||
```

```
Db      15873 CGCTCTGTGTCTAGCTAAAGGTTTGTAAACGCACCAATCAGCACTCTGTAAAAACAGACC 15814

Qy        543 ----------------------------------------ATGGACCAATCAG------ 555
                                                       ||||||||||||
Db      15813 AATCAGCACTCTGTAAAACGGACCAATCAGCACTCTGTAAAATGGACCAATCAGCTCTCT 15754

Qy        556 ---------------CAGGATGTGGGCGGGGTCAAATAAGGGAGT--------AAAAACT 592
                            | | ||||||| |||| |||||||||| |         |   |
Db      15753 GTAAAATGGACCAATCTGCATGTGGGTGGGGCCAAATAAGGGATTTCCCCCCCCCACCTCC 15694

Qy        593 GGCCACCCGAGCCAGCAGTGGCAACCCACTCGGGTCCCCTTCCACACTGTGGAAG-CTTT 651
              | ||  || |  ||||||| ||||||  ||| |||   ||||||  |||||||||  | |
Db      15693 GCCCTTCCCAAGCAGCAGTAGCAACCTGCTCAGGTTGTCTTCCAGGCTGTGGAAGTGTAT 15634

Qy        652 GTTCTTTTGCTCTTCACAATAAATCTTGCTGCTGCTCATTCTTTGTGTCCACACTACCTT 711
              ||||||  ||| |||| ||||||| ||||| |||||| |||||| ||    |||||||||
Db      15633 TTTCTTTCACTCGTCACCATAAATCCTGCTGTTGCTCACTCTTTGGGTTTGAACTACCTT 15574

Qy        712 TATGAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCATTCCTGAAGTCA-ACAGACCA 770
              |||||||| ||||| |||||  || ||||||  ||||||  |||||||| ||  |||||||
Db      15573 TATGAGCTATAACATTCACTATGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCCAGACCA 15514

Qy        771 CGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCCTTTAAGAGCTGTAACACTC 830
              ||||||||    ||||| ||| |||||| || | ||  ||||||||||||||| |||||
Db      15513 TGAACCCACCAAGAGGAATAAACAACTCCAGACGCGCCACCTTTAAGAGCTGTAGCACTC 15454

Qy        831 ACTGCGAAGCTCTGCAGCTTCACTCCTGAAGTCAGTGAGACCACAAACCCACCAGAAGGA 890
              |||||||||  |||||  |||||||  ||||||||||  |||||||||||||||  ||||||
Db      15453 ACTGCGAAGGTCTGCGGCTTCACCCCTGAAGTCAGCAAGACCACAAACCCACCGGAAGGA 15394

Qy        891 AGAAACTCTGGACACACCTGAATATCTGAAGGAACAAACTCCAGACACACCATCTTTCAG 950
              |||||||  |||||||  |||||  |||||||||||||||||||  ||||||||||||| ||
Db      15393 AGAAACTGCGGACACATCTGAACATCTGAAGGAACAAACTCCGGACACACCATCTTTAAG 15334

Qy        951 AGCTGTAACACTCACCGCAAGGGTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAA 1010
              | ||||||||||||||||    || | | ||||||||||||  ||||||||||||||||||
Db      15333 AACTGTAACACTCACCGTGTAGGACCGCGGCTTCATTCTTAGAGTCAGCAAGACCAAGAA 15274

Qy       1011 CCCACCGGAAGGAACAAATTCCAGACACA 1039
              | |||| |||||||  ||||  ||||||
Db      15273 CTCACCAGAAGGAATCAATTTTGGACACA 15245
```

Search completed: January 22, 2006, 21:31:03
Job time : 275.224 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:10:03 ; Search time 5419.54 Seconds
(without alignments)
17274.699 Million cell updates/sec

Title: US-09-728-552A-3_COPY_78595_80595
Perfect score: 2001
Sequence: 1 atctttatttgtatgagaaa.....aagtcattctgccagaattc 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	650.4	32.5	652	9	B72263	B72263 RPCI11-7O22
c 2	478.8	23.9	517	9	AQ251771	AQ251771 HS_3202_B
3	453	22.6	466	9	AQ737783	AQ737783 HS_2250_B
c 4	423	21.1	434	9	AQ593093	AQ593093 HS_5457_A
c 5	346	17.3	359	9	AQ822612	AQ822612 HS_5558_A
6	249	12.4	660	9	AQ353839	AQ353839 RPCI11-10
c 7	165	8.2	792	10	AG540684	AG540684 Mus muscu
c 8	163.2	8.2	597	9	AZ706600	AZ706600 RPCI-23-2
c 9	144	7.2	375	9	AZ065413	AZ065413 RPCI-23-3
10	140.8	7.0	544	9	AQ463585	AQ463585 HS_5J12_A
11	121.8	6.1	707	10	AG310514	AG310514 Mus muscu
12	72.6	3.6	550	9	AZ879366	AZ879366 RPCI-23-1
c 13	67	3.3	822	10	AG282114	AG282114 Mus muscu
14	57.4	2.9	905	10	CNS00KHX	AL077798 Drosophil
c 15	53.8	2.7	734	9	AQ478276	AQ478276 RPCI-11-2
16	53.2	2.7	996	10	CNS00FUH	AL071063 Drosophil
17	51.8	2.6	997	10	CNS005TE	AL060767 Drosophil
18	51.8	2.6	1200	10	CNS016CO	AL106578 Drosophil
19	51	2.5	987	10	CNS014PQ	AL104456 Drosophil
c 20	50	2.5	1101	10	CNS0039G	AL063921 Drosophil
c 21	48.8	2.4	377	9	AQ212826	AQ212826 HS_3118_B
22	48.8	2.4	1062	1	AJ925404	AJ925404 AJ925404
c 23	48.8	2.4	1254	10	CL105409	CL105409 ISB1-44M8
c 24	48.2	2.4	391	9	AQ338558	AQ338558 HS_3118_B
c 25	48.2	2.4	1204	10	CNS016E2	AL106628 Drosophil
26	48	2.4	958	10	CW951690	CW951690 TcB37.2_A
27	48	2.4	1092	10	CNS020K7	AL175696 Tetraodon
28	47.8	2.4	950	1	AL551397	AL551397 AL551397
29	47.6	2.4	1140	10	CL071552	CL071552 CH216-121
c 30	47.4	2.4	548	1	AU062068	AU062068 AU062068
c 31	47.4	2.4	1201	10	CNS015YI	AL106068 Drosophil
32	47.2	2.4	1225	10	CNS0161D	AL106171 Drosophil
33	46.8	2.3	1281	10	CL022041	CL022041 CH216-13E
34	46.6	2.3	511	8	DR071573	DR071573 RTDK1_20_
c 35	46.4	2.3	617	7	CV569142	CV569142 od01h09.y
36	46.4	2.3	644	9	BH712636	BH712636 BOMFC43TF
37	46.4	2.3	649	9	AZ899171	AZ899171 RPCI-24-2
38	46.4	2.3	733	9	AZ899889	AZ899889 RPCI-24-2
39	46.4	2.3	1001	10	CNS0155H	AL105023 Drosophil
40	46.2	2.3	1364	10	CG757986	CG757986 P053-2-D0
c 41	46	2.3	778	8	CV784532	CV784532 UI-M-HY0p
c 42	46	2.3	889	8	DN573331	DN573331 92944934
c 43	46	2.3	928	10	CNS00DKY	AL071865 Drosophil
c 44	46	2.3	1201	10	CNS01038	AL098462 Drosophil
c 45	46	2.3	1225	10	CNS0161D	AL106171 Drosophil
46	45.8	2.3	669	9	BH996047	BH996047 oeh20a09.
47	45.8	2.3	747	5	BW167551	BW167551 BW167551
c 48	45.8	2.3	792	9	BZ466334	BZ466334 BONRX04TF
c 49	45.8	2.3	928	10	CZ505964	CZ505964 GMW2-21N2
50	45.8	2.3	1201	10	CNS016EA	AL106636 Drosophil
51	45.6	2.3	1341	10	CG755648	CG755648 P051-1-G0
c 52	45.4	2.3	363	2	BF083948	BF083948 CM4-CT048
53	45.4	2.3	1101	10	CNS00EVL	AL069706 Drosophil
54	45.4	2.3	1101	10	CNS00LT2	AL078714 Drosophil
c 55	45.4	2.3	1201	10	CNS0168C	AL106422 Drosophil
c 56	45.2	2.3	738	5	BU929426	BU929426 AGENCOURT
c 57	45.2	2.3	1101	10	CNS0100X	AL098379 Drosophil
58	45	2.2	1101	10	CNS0039G	AL063921 Drosophil
59	45	2.2	1101	10	CNS00EPO	AL069493 Drosophil
60	45	2.2	1134	1	AL563015	AL563015 AL563015
61	45	2.2	1214	10	AG436204	AG436204 Mus muscu
c 62	44.8	2.2	337	2	BF449162	BF449162 maa43g01.
c 63	44.8	2.2	578	6	CF532184	CF532184 UI-M-GH0-
c 64	44.8	2.2	584	6	CF748442	CF748442 UI-M-HJ0-
c 65	44.8	2.2	606	7	CK782134	CK782134 UI-M-HJ0-
c 66	44.8	2.2	718	9	BH965542	BH965542 odj15a10.
c 67	44.8	2.2	727	9	BZ037257	BZ037257 oef92f11.
c 68	44.8	2.2	794	8	CX240391	CX240391 NMA03060
c 69	44.8	2.2	807	3	BQ178571	BQ178571 UI-M-EV0-
c 70	44.8	2.2	941	10	CW939056	CW939056 TcB19.2_C
c 71	44.8	2.2	1101	10	CNS016LI	AL106896 Drosophil
72	44.6	2.2	512	2	BF198281	BF198281 248042 MA
c 73	44.6	2.2	800	10	CNS02I1J	AL198352 Tetraodon
74	44.6	2.2	815	8	CX588572	CX588572 TTE000323
75	44.6	2.2	928	10	CNS00DKY	AL071865 Drosophil
c 76	44.6	2.2	1036	10	CNS010BS	AL098770 Drosophil
77	44.6	2.2	1043	10	CNS0145P	AL103735 Drosophil
c 78	44.6	2.2	1043	10	CNS0145P	AL103735 Drosophil
c 79	44.6	2.2	1113	10	AG390738	AG390738 Mus muscu
c 80	44.4	2.2	238	11	CNS04CWO	AL285009 Tetraodon
81	44.4	2.2	627	3	BM164497	BM164497 EST567020
c 82	44.4	2.2	1269	8	DN809239	DN809239 76956677
83	44.2	2.2	371	9	AZ620815	AZ620815 1M0453M07
c 84	44.2	2.2	681	9	BH008726	BH008726 ef19b02.x
c 85	44.2	2.2	689	9	BH999767	BH999767 oed95g06.
86	44.2	2.2	795	10	AG044101	AG044101 Pan trogl
c 87	44.2	2.2	855	10	CZ545757	CZ545757 SRAA-aad6
c 88	44.2	2.2	912	9	AZ545767	AZ545767 ENTGC39TR
89	44.2	2.2	964	10	CNS006N9	AL065781 Drosophil
90	44.2	2.2	1004	10	CNS00BV2	AL058076 Drosophil
c 91	44.2	2.2	1101	10	CNS00ETW	AL069847 Drosophil
92	44.2	2.2	1189	1	AJ925925	AJ925925 AJ925925
c 93	44.2	2.2	1203	10	CNS015WU	AL106008 Drosophil
c 94	44	2.2	502	2	BF937276	BF937276 fm62h05.y
c 95	44	2.2	538	2	BG180961	BG180961 602327440

c 96	44	2.2	550	6	CA848973	CA848973	ir68e04.y
97	44	2.2	618	5	BQ632119	BQ632119	il22f11.x
c 98	44	2.2	699	2	BG180801	BG180801	602327232
c 99	44	2.2	742	5	BU852860	BU852860	AGENCOURT
100	44	2.2	775	5	BU626054	BU626054	UI-H-FG1-

ALIGNMENTS

RESULT 1
B72263/c
LOCUS B72263 652 bp DNA linear GSS 08-APR-1999
DEFINITION RPCI11-7O22.TV RPCI-11 Homo sapiens genomic clone RPCI-11-7O22, genomic survey sequence.
ACCESSION B72263
VERSION B72263.1 GI:2711414
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 652)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI11-7O22.TP
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .652
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7502661"
/db_xref="taxon:9606"
/clone="RPCI-11-7O22"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
ORIGIN

Query Match 32.5%; Score 650.4; DB 9; Length 652;
Best Local Similarity 99.8%; Pred. No. 4.5e-157;
Matches 651; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1070 GTGGAACTGACCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTC 1129
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 GTGGAACTGACCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTC 593

Qy       1130 TGCTGACTCTCCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAG 1189
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         592 TGCTGACTCTCCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAG 533

Qy       1190 TGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGT 1249
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         532 TGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGT 473

Qy       1250 GAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTA 1309
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         472 GAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTA 413

Qy       1310 ATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCAT 1369
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         412 ATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCAT 353

Qy       1370 CTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAG 1429
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         352 CTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAG 293

Qy       1430 CTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAG 1489
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         292 CTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAG 233

Qy       1490 GGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTC 1549
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         232 GGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTC 173

Qy       1550 ATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGT 1609
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         172 ATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGT 113

Qy       1610 CATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTC 1669
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         112 CATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTC 53

Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAAT 1721
             ||||||||||||||||||||||||||||||||||||||||| |||||||||
Db          52 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATAGTTGTTTAAT 1
```

RESULT 2
AQ251771/c
LOCUS AQ251771 517 bp DNA linear GSS 10-OCT-1998
DEFINITION HS_3202_B1_G12_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3202 Col=23 Row=N, genomic survey sequence.
ACCESSION AQ251771
VERSION AQ251771.1 GI:3722344
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 517)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Purlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3202 row: N column: 23
Class: BAC ends
High quality sequence stop: 517.
FEATURES Location/Qualifiers
source 1. .517
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```
                     /clone="Plate=3202 Col=23 Row=N"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

  Query Match             23.9%;  Score 478.8;  DB 9;  Length 517;
  Best Local Similarity   97.1%;  Pred. No. 1.3e-112;
  Matches  497;  Conservative    0;  Mismatches   14;  Indels    1;  Gaps    1;

Qy        827 ATTGATTCTTTTTGACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCT 886
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 AATGATTCTTTTTGACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCT 455

Qy        887 ATGGAAACCAACACGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGA 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 ATGGAAACCAACACGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGA 395

Qy        947 TGTTCTAGAGAGCTGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACT 1006
              |||| |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        394 TGTTATAGAGAGCTGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCGCAGCACT 335

Qy       1007 GAAATTAACACCACTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCA 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 GAAATTAACACCACTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCA 275

Qy       1067 GGAGTGGAACTGACCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTT 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 GGAGTGGAACTGACCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTT 215

Qy       1127 TTCTGCTGACTCTCCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGT 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 TTCTGCTGACTCTCCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGT 155

Qy       1187 AAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACAC 1246
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        154 AAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACNC 95

Qy       1247 CGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAG-CCC 1305
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||| |||
Db          94 CGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTGTCACAAGCCCC 35

Qy       1306 CCTAATTTTACTGCATAATTATTTTGAATTCA 1337
              ||||||| | |||||||||     ||  ||||
Db          34 CCTAATTNTGCTGCATAATATAGATGGTTTCA 3


RESULT 3
AQ737783
LOCUS       AQ737783                 466 bp    DNA     linear   GSS 16-JUL-1999
DEFINITION  HS_2250_B2_E02_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2250 Col=4 Row=J, genomic survey
            sequence.
ACCESSION   AQ737783
VERSION     AQ737783.1  GI:5515305
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2250  row: J  column: 4
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 466.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2250 Col=4 Row=J"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

  Query Match             22.6%;  Score 453;  DB 9;  Length 466;
  Best Local Similarity   99.6%;  Pred. No. 6.3e-106;
  Matches  464;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy        522 CTGCAGGGTTTCTACTAAGCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCA 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGCAGGGTTTCTACTAAGCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCA 60

Qy        582 CTCTAGTGATCTTTCTCTTTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCC 641
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 CTCTAGTGATCTTTCTCTTTTCCTCACCT-TGGGACTGGTGGCTGTTTGTATGGACTGCC 119

Qy        642 TTAGCTTTGCTTTGGGTTTTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATA 701
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        120 TTAGCTTTGCTTTGGGTTTTTTCCTGNGGACAATGTCTTCAGATTATCCTAGACCAAATA 179

Qy        702 AACTACAGCCACTGGGCCAGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTT 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 AACTACAGCCACTGGGCCAGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTT 239

Qy        762 CACCTTAGTTTAGGTCAAGCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTC 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 CACCTTAGTTTAGGTCAAGCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTC 299

Qy        822 TAGCAATTGATTCTTTTTGACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATT 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 TAGCAATTGATTCTTTTTGACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATT 359

Qy        882 CCCCTATGGAAACCAACACGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATA 941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 CCCCTATGGAAACCAACACGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATA 419

Qy        942 CAGGATGTTCTAGAGAGCTGTATTTTGGGCACTTAACTATTCTCCA 987
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        420 CAGGATGTTCTAGAGAGCTGTATTTTGGGCACTTAACTATTCTCCA 465


RESULT 4
AQ593093/c
LOCUS       AQ593093                 434 bp    DNA     linear   GSS 08-JUN-1999
DEFINITION  HS_5457_A2_F10_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1033 Col=20 Row=K, genomic survey sequence.
ACCESSION   AQ593093
VERSION     AQ593093.1  GI:5024745
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
```

```
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1033  row: K  column: 20
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=1033 Col=20 Row=K"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             21.1%;  Score 423;  DB 9;  Length 434;
  Best Local Similarity   98.4%;  Pred. No. 3.7e-98;
  Matches  426;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1110 CAGAAGTCAGCTGTGTTTTCTGCTGACTCTCCATAGGAACATCAGTCCTGAATCCTCAGA 1169
             ||||||||||||||| || |||||||||||||||||||||||||| ||||||||||||
Db        433 CAGAAGTCAGCTGTGTNTTGTGCTGACTCTCCATAGGAACATCAGTCATGAATCCTCAGA 374

Qy       1170 CCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGGATCTC 1229
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 CCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGGATCTC 314

Qy       1230 CAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAATCTT 1289
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 CANAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAATCTT 254

Qy       1290 GCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCT 1349
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 GCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCT 194

Qy       1350 ACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTA 1409
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 ACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTA 134

Qy       1410 ATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGC 1469
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 ATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGC 74

Qy       1470 TCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAG 1529
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          73 TCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAG 14

Qy       1530 GTCTCATATGTAA 1542
             ||||||||   ||
Db         13 GTCTCATAAAAAA 1


RESULT 5
AQ822612/c
LOCUS       AQ822612                359 bp    DNA     linear   GSS 26-AUG-1999
DEFINITION  HS_5558_A1_F10_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1134 Col=19 Row=K, genomic survey sequence.
ACCESSION   AQ822612
VERSION     AQ822612.1  GI:5785005
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1134  row: K  column: 19
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 359.
FEATURES             Location/Qualifiers
     source          1. .359
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=1134 Col=19 Row=K"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             17.3%;  Score 346;  DB 9;  Length 359;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-78;
  Matches  346;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1413 CATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCT 1472
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 CATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCT 300

Qy       1473 TGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTC 1532
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 TGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTC 240

Qy       1533 TCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTC 1592
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTC 180

Qy       1593 ACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGA 1652
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 ACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGA 120

Qy       1653 ACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATAT 1712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATAT 60

Qy       1713 TTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTC 1758
              |||||||||||||||||||||||||||||||||||||||||||||
Db         59 TTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTC 14


RESULT 6
AQ353839
LOCUS       AQ353839                660 bp    DNA     linear   GSS 07-MAY-1999
DEFINITION  RPCI11-109P10.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-109P10,
            genomic survey sequence.
ACCESSION   AQ353839
VERSION     AQ353839.1  GI:4176797
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-109P10.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7541841"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-109P10"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             12.4%;  Score 249;  DB 9;  Length 660;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-53;
  Matches  249;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1753 GAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTG 1812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTG 64

Qy       1813 TGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTACAGGTGCCAAGTT 1872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 TGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTACAGGTGCCAAGTT 124

Qy       1873 AAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTTAGTGAACTG 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         125 AAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTTAGTGAACTG 184

Qy       1933 AAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGTGAAGTCATTCTG 1992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         185 AAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGTGAAGTCATTCTG 244

Qy       1993 CCAGAATTC 2001
              |||||||||
Db         245 CCAGAATTC 253


RESULT 7
AG540684/c
LOCUS       AG540684                792 bp    DNA     linear   GSS 23-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-453L21.TJ, genomic survey
            sequence.
ACCESSION   AG540684
VERSION     AG540684.1  GI:48301098
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 792)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .792
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-453L21.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match              8.2%;  Score 165;  DB 10;  Length 792;
  Best Local Similarity   63.8%;  Pred. No. 3.5e-31;
  Matches  340;  Conservative    0;  Mismatches  170;  Indels   23;  Gaps    5;

Qy       1044 TAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGACCTCTTATGTGTGTCCAGCCTTTC 1103
```

```
                 || ||    |   | ||| ||||| || ||||   ||| ||| |||||| |||||||| |
Db        590 TACTCATGTGTGATACAGTCTCAGAAGAGGAAGCCACCACTTCTGTGTGGCCAGCCTTCC 531

Qy       1104 TTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTCCATAGGAACATCAGT-CCTGAAT 1162
                ||    |      |   ||  |||     ||  |  ||  |||||   ||| ||||||
Db        530 CTCTGAAATGGCCCTCTTGGCTTTCTATGTGGATGCCCTCAGGAATGCAAGTACCTGAAT 471

Qy       1163 CCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCT 1222
                || || ||||||  ||| |||| ||||   ||       ||     |         || |
Db        470 CCCCAAACCACCTCCTGAAGTACTAAGCATTCTGTCAGTTCAGCAAGCCCCCACCCCCCG 411

Qy       1223 GGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGT 1282
                ||| | |||   | ||||    ||  ||| |||| |||||||| | |  ||| ||||| |
Db        410 GGAGCCCCAGTTCCTGTGGTGAACTATGATAGAGGAATGAGAAGGGTCAGCTTTTCAGAT 351

Qy       1283 AAA-------------TCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTATTT 1329
                |||             | |||||||| |  |  ||| || ||| |||||||| ||  ||
Db        350 AAACTTTGCCTCTTTGTTTTGCTTTTTAAACAAGCCCATAGTTTCACTGCATAGTTGCTT 291

Qy       1330 TGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCA 1389
                ||||| ||||||||||||    |||||| ||  |||||||  ||| | | | |||||| |
Db        290 TGAATCCACTGATAATTTGGGCAATTTCCCTGTAAGTCATTCACAAATAGTTCCCTCTAA 231

Qy       1390 TGCAACACTTGGCTTTGCTAATACAT------ATCTATTATGAGAGCTGTGCTTCTTAAG 1443
                ||    | ||  ||| |||     || |          |  ||||||||||||||| ||
Db        230 TGTTGCCCTCAGCTATGCGGCCACGTTATCTGTCTATTACTGAGAGCTGTGCTTCTTGAG 171

Qy       1444 CGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAAT 1503
                |||||||||||| |||| |||||||||||||||||||  |||||  ||||  |||
Db        170 TATAAATGTTTTATGTGCATTAAGGCTCTTGGCTTACATAT-CAAGGGACATTGGACAAC 112

Qy       1504 GTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGAT 1556
                |  |||||||| |||| |   ||  | ||| | ||||||||||||    |||
Db        111 GGTATACAGAAGCCTTTGCGTGAC--GACTCGTGTGTAAAGAATTCCGCGGAT 61


RESULT 8
AZ706600/c
LOCUS       AZ706600                597 bp    DNA     linear   GSS 24-JAN-2001
DEFINITION  RPCI-23-209G10.TJ RPCI-23 Mus musculus genomic clone
            RPCI-23-209G10, genomic survey sequence.
ACCESSION   AZ706600
VERSION     AZ706600.1  GI:12434225
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
            Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-209G10.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 209  row: G  column: 10
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-209G10"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
ORIGIN

  Query Match             8.2%;  Score 163.2;  DB 9;  Length 597;
  Best Local Similarity   65.5%;  Pred. No. 9.4e-31;
  Matches  325;  Conservative    0;  Mismatches  148;  Indels   23;  Gaps    5;

Qy       1044 TAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGACCTCTTATGTGTGTCCAGCCTTTC 1103
                || ||    |   | ||| ||||| || ||||   ||| ||| |||||| |||||| | |
Db        494 TACTCATGTGTGATACAGTCTCAGAAGAGGAAGCCACCACTTCTGTGTGGCCAGCCCTCC 435

Qy       1104 TTCCTTCAGAAGTCAGCTGTGTTTTC--TGCTGACTCTCCATAGGAACATCAGTCCTGAA 1161
                  ||  |      |  ||  | ||||  ||  ||  | |    | |        ||||||
Db        434 CCTCTGAAATGACCCTCTTGGCTTTCTATGTGGATGCCCTCAGGAATGCAAGTACCTGAA 375

Qy       1162 TCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGC 1221
                ||| || ||||||  ||| |||| ||||   | ||| ||||  |   |||    |  || |
Db        374 TCCCCAAACCACCTCCTGAAGTACTAAGCATT-CTGTCAGTTCAGCAAGCCCCCACCCCC 316

Qy       1222 TGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGG 1281
                ||| | |||   | ||||    ||  |||||||||||||||| | |  ||| |||||
Db        315 GGGAGCCCCAGTTCCTGTGGTGAACTATGAGAGAGAAATGAGAAGGGTCAGCTTTTCAGA 256

Qy       1282 TAAA-------------TCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTATT 1328
                ||||             | |||||||| |  |  ||| || ||| |||||||| ||  |
Db        255 TAAACTTTGCCTCTTTGTTTTGCTTTTTAAACAAGCCCATAGTTTCACTGCATAGTTGCT 196

Qy       1329 TTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTC 1388
                |||||| ||||||||||||    |||||| ||  ||||||||  ||| | | | ||||||
Db        195 TTGAATCCACTGATAATTTGGGCAATTTCCCTGTAAGTCATTCACAAATAGTTCCCTCTG 136

Qy       1389 ATGCAACACTTGGCTTTGCTAATACATATCT-----ATTATGAGAGCTGTGCTTCTTAAG 1443
                |||   | ||  ||| |||     || |||||       |  ||||||||||||||| ||
Db        135 ATGTTGCCCTCAGCTATGCGGCCACGTATCTGTCTATTACTGAGAGCTGTGCTTCTTGAG 76

Qy       1444 CGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAAT 1503
                |||||||||||| |||| ||||||||||||||||||||||  ||  |||   ||  |||
Db         75 TATAAATGTTTTATGTGCATTAAGGCTCTTGGCTTACATATCAA--GGGGCATGGACAAC 18

Qy       1504 GTGATACAGAAGTCTT 1519
                | |||||||||| |||
Db         17 GAGATACAGAAGCCTT 2


RESULT 9
AZ065413/c
LOCUS       AZ065413                375 bp    DNA     linear   GSS 30-MAR-2000
DEFINITION  RPCI-23-391K3.TVB RPCI-23 Mus musculus genomic clone RPCI-23-391K3,
            genomic survey sequence.
ACCESSION   AZ065413
VERSION     AZ065413.1  GI:7356665
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
```

```
            Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-391K3.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 391  row: K  column: 3
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-391K3"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
ORIGIN

  Query Match             7.2%;  Score 144;  DB 9;  Length 375;
  Best Local Similarity   69.4%;  Pred. No. 7.9e-26;
  Matches  261;  Conservative    0;  Mismatches   95;  Indels   20;  Gaps    4;

Qy       1161 ATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCG 1220
             |||| || ||||||  ||| |||| ||||    | ||| |||| |   |||     |  ||
Db        375 ATCCCCAAACCACCTCCTGAAGTACTAAG-CATTCTGTCAGTTCAGCAAGCCCCCACCCC 317

Qy       1221 CTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAG 1280
             | ||| | |||   | ||||    ||  |||||||||||||||| | |  ||| |||||
Db        316 CGGGAGCCCCAGTTCCTGTGGTGAACTATGAGAGAGAAATGAGAAGGGTCAGCTTTTCAG 257

Qy       1281 GTAAA------------TCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTAT 1327
              ||||            | |||||||||  |  |  ||| || ||| ||||||||  ||
Db        256 ATAAACTTTGCCTCTTTGTTTTGCTTTTTAAACAAGCCCATAGTTTCACTGCATAGTTGC 197

Qy       1328 TTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCT 1387
             |||||||  |||||||||  |   ||||||  ||  ||||||||  ||| | | | ||||||
Db        196 TTTGAATCCACTGATAAAATGGGCAATTTCCCTGTAAGTCATTCACAAATAGTTCCCTCT 137

Qy       1388 CATGCAACACTTGGCTTTGCTAATACATATCT-----ATTATGAGAGCTGTGCTTCTTAA 1442
              |||   | ||  ||| |||    || |||||      |  |||||||||||||||| |
Db        136 GATGTTGCCCTCAGCTATGCGGCCACGTATCTGTCTATTACTGAGAGCTGTGCTTCTTGA 77

Qy       1443 GCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAA 1502
              |  |||||||||||| |||| ||||||||||||||||||||  ||||||  ||||  |||
Db          76 GTATAAATGTTTTATGTGCATTAAGGCTCTTGGCTTACATAT-CAAGGGGCATTGGACAA 18

Qy       1503 TGTGATACAGAAGTCT 1518
              | ||||||||||| ||
Db          17 CGAGATACAGAAGCCT 2
```

```
RESULT 10
AQ463585
LOCUS       AQ463585                544 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  HS_5212_A1_E01_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=788 Col=1 Row=I, genomic survey sequence.
ACCESSION   AQ463585
VERSION     AQ463585.1  GI:4640680
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 788  row: I  column: 1
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 544.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=788 Col=1 Row=I"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             7.0%;  Score 140.8;  DB 9;  Length 544;
  Best Local Similarity   74.9%;  Pred. No. 5.9e-25;
  Matches  191;  Conservative    0;  Mismatches   58;  Indels    6;  Gaps    1;

Qy       1751 TTGAATTCTTTATGATGAGAACTATCTTGACTC------AAGTGGACAGTGGTGAGCTTT 1804
              ||| |||||||||| ||||||| |  |  |       |||  ||   |     | |
Db        129 TTGGATTCTTTATGACGAGAACTCTTCTTTGACCTCAAGCAGTACACGTGGTGAGCTTCT 188

Qy       1805 TTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTACAGGT 1864
              ||||||||||||||||    |||||| |||||||||| |  ||||||||  ||| ||
Db        189 CTGGCCTGTGGTCCCTACCCTAAAAGGAAGCTTTGTCATACATCCTTATATGAGACATGT 248

Qy       1865 GCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTTA 1924
              |||||  |||  ||| || ||||||||||||||||||||||||  ||||  ||||||||||
Db        249 GCCAACATAATGGCCTAACCTTGCTCTTAAAAGCATACTGGAATTTGGGTTAGACTTTTA 308

Qy       1925 GTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGTGAAG 1984
              |||||   | |||||||||| || ||| ||| ||||||||||| |||||||| ||||
Db        309 CTGAACCTGAAGGAATAAACACATGCCTGTGGCAGAACTTCTCCTTCATCCTTGGAGAAG 368
```

```
Qy       1985 TCATTCTGCCAGAAT 1999
              ||||||| ||  | |
Db        369 TCATTCTCCCNCACT 383


RESULT 11
AG310514
LOCUS       AG310514                 707 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-092P04.TJ, genomic survey
            sequence.
ACCESSION   AG310514
VERSION     AG310514.1  GI:47883468
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 707)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-092P04.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             6.1%;  Score 121.8;  DB 10;  Length 707;
  Best Local Similarity   63.2%;  Pred. No. 5.4e-20;
  Matches  265;  Conservative    0;  Mismatches  127;  Indels   27;  Gaps    4;

Qy       1544 GAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAA 1603
              ||||||| | | |||| |||||||||||||||||||||||||  |||||||||     ||
Db         95 GAATTCACTGGGTTGGGTGAAATAGACTGATCTGTCCATTTCCGTGCTCACTTCCAGCAA 154

Qy       1604 GGAAGTCATTA----GCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCT 1659
              |||||||| ||     | | | | | |||  |||  |||||| ||||| ||||||| | |
Db        155 GGAAGTCACTACCAGCCAAGGAACCGAAAGGTACTGTCTATGCAATTAAAAGAACAGGTT 214

Qy       1660 GGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTA 1719
              || |||||||||| | ||     | |||||||||||  | || || | | ||| ||||
Db        215 GGCTTTGCTCAATACAGAA-----ACAAAGAAACCATGCAAGAGACACAGTGTTTCTTTA 269

Qy       1720 ATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTAT---- 1775
              || |||  | ||||| |||||| || |   |||    |  |||| ||     || |
Db        270 ATTAGGAGACTGAGATTCTATTGAAGACACTTTATGATTATTATTATTTTTTCTTTATGG 329

Qy       1776 ----------CTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGT 1824
                      |||||||||||||  |||| ||  || |   | || | |  |||| |
Db        330 TGGCAGCTATCCTTGACTCAAGTGGGAAGTGCTGGCCTCT---GGCTTTTGCACCTATGC 386

Qy       1825 AGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGC 1884
              ||| | | || ||||  |||||  || | ||||| | ||| |  |  | | |  | |||
Db        387 AGAGAAGTGGTTTTGCTATAAACGCTCAAATGGTGCCGGTTCGGAACTCAAGGTTCCAGC 446

Qy       1885 TTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTTAGTGAACTGAAGGGAATAAA 1943
              |||||| ||| |   |    | ||||   |         ||      | |||| | ||||
Db        447 TTGCTCCTAACATTCTGGATGGTTTTTAATCCAGGAACTACCTGTTTCAAGGCATTAAA 505


RESULT 12
AZ879366
LOCUS       AZ879366                 550 bp    DNA     linear   GSS 05-MAR-2001
DEFINITION  RPCI-23-185B21.TJ RPCI-23 Mus musculus genomic clone
            RPCI-23-185B21, genomic survey sequence.
ACCESSION   AZ879366
VERSION     AZ879366.1  GI:13197432
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
            Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-185B21.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 185  row: B  column: 21
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-185B21"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
ORIGIN
```

```
  Query Match             3.6%;  Score 72.6;  DB 9;  Length 550;
  Best Local Similarity   60.9%;  Pred. No. 2.9e-07;
  Matches  199;  Conservative    0;  Mismatches  104;  Indels   24;  Gaps    4;

Qy       1632 ACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAA 1691
             ||  |||||| ||||| ||||||| | ||| |||||      ||| |   | | ||||||
Db          1 ACTGTCTATGCAATTAAAAGAACAGGTTGGCTTTGC------TAATACAGAAACAAAGAA 54

Qy       1692 ACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATT 1751
             ||||||  | || || | | ||| |||||| |||  | ||||| ||||||  | |   ||
Db         55 ACCATGCAAGAGACACAGTGTTTCTTTAATTAGGAGACTGAGATTCTATTGTAGACACTT 114

Qy       1752 TGAA--------------TTCTTTATGATGAGAACTAT-CTTGACTCAAGTGGACAGTGG 1796
              |                |||||||| ||  | |||| |||||||||||||  ||||
Db        115 TATGATTATTATTTTTTTTCTTTATGGTGGCAGCTATCCTTGACTCAAGTGGGAAGTGT 174

Qy       1797 TGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATG 1856
             ||  || |   | || | |  |||| | ||| | | || ||||  |||||  || | |||
Db        175 TGGCCTCT---GGCTTTTGCACCTATGCAGAGAAGTGGTTTTGCTATAAACGCTCAAATG 231

Qy       1857 GTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTA 1916
             || | ||| |  |  || | | |||||||||| ||| |   |     | ||||   |
Db        232 GTGCCGGTTCGGAACTCAGGGTTCCAGCTTGCTCCTAACATTCTGGATGGTTTTTAATCC 291

Qy       1917 GACTTTTAGTGAACTGAAGGGAATAAA 1943
                 ||      | |||| | ||||
Db        292 AGGAACTACCTGTTTCAAGGCATTAAA 318


RESULT 13
AG282114/c
LOCUS       AG282114                 822 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-054B06.TJ, genomic survey
            sequence.
ACCESSION   AG282114
VERSION     AG282114.1  GI:47854991
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 822)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .822
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-054B06.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             3.3%;  Score 67;  DB 10;  Length 822;
  Best Local Similarity   78.9%;  Pred. No. 9.1e-06;
  Matches  105;  Conservative    0;  Mismatches   25;  Indels    3;  Gaps    2;

Qy       1424 TGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATA 1483
             ||||||||||||||||  ||  |||||||||||| |||| ||||||||||||||||||||
Db        194 TGAGAGCTGTGCTTCTTGAGTATAAATGTTTTATGTGCATTAAGGCTCTTGGCTTACATA 135

Qy       1484 TAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAA 1543
             |  |||||  ||||  ||| |  ||||||||| |||| |    ||  | ||| | ||||||
Db        134 T-CAAGGGACATTGGACAACGGTATACAGAAGCCTTTGCGTGAC--GACTCGTGTGTAAA 78

Qy       1544 GAATTCATTAGAT 1556
                ||||||    |||
Db         77 GAATTCCGCGGAT 65


RESULT 14
CNS00KHX
LOCUS       CNS00KHX                 905 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR17N06 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL077798
VERSION     AL077798.1  GI:4957174
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 905)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .905
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR17N06"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN
```

```
  Query Match             2.9%;  Score 57.4;  DB 10;  Length 905;
  Best Local Similarity   36.5%;  Pred. No. 0.0029;
  Matches  105;  Conservative   51;  Mismatches  132;  Indels    0;  Gaps    0;

Qy       1493 TATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATT 1552
Db        493 TNNTTGNCBWTGTTMTTTTTAWTTGKTKTTKATAAATGTTWGKKKKKRRWTWKTTKDKTK 552

Qy       1553 AGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCAT 1612
Db        553 AAAATTATWAAKAWAAAKATKWKKTWTTDTKWWKWTAWKAAAAAAHAAWATAAATWATAA 612

Qy       1613 TAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAAT 1672
Db        613 TAWATAAAAAAAAATAAATWAAAAAAAAAAAAAATAAAWWATAAAAAATATWTWTTAWAAA 672

Qy       1673 ATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGA 1732
Db        673 AAAAAAAAAWAAAWAWAAAAAAAAAAAAAAAAAAAAAAAATAAAAWATAWWTTAWWTTWWWTA 732

Qy       1733 GAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGA 1780
Db        733 WWAWTAWAWAAWAAAAATAATAWWAAWTAAWARWGATAWATATATWAA 780


RESULT 15
AQ478276/c
LOCUS       AQ478276                734 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  RPCI-11-254C3.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-254C3,
            genomic survey sequence.
ACCESSION   AQ478276
VERSION     AQ478276.1  GI:4660395
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-254C3.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7597202"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-254C3"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             2.7%;  Score 53.8;  DB 9;  Length 734;
  Best Local Similarity   62.0%;  Pred. No. 0.023;
  Matches   85;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          4 TTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCA 63
Db        567 TTCTTTTGTCTGATTATAAACAAAGGATTGCTAATAATTCACTGAAATGTATTAGTTGTG 508

Qy         64 GGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAAT 123
Db        507 GATATGTTAATTGCAAGTGACAGAAACCCTACTCAAACTGGGTTAAGCAAAATGGGGAAT 448

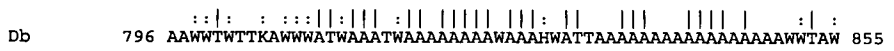
Qy        124 ATATTGGCTCATGTAAC 140
Db        447 TTATTGGTTCATAAAAC 431


RESULT 16
CNS00FUH
LOCUS       CNS00FUH                996 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR31O21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071063
VERSION     AL071063.1  GI:4951105
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 996)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .996
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR31O21"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             2.7%;  Score 53.2;  DB 10;  Length 996;
  Best Local Similarity   45.1%;  Pred. No. 0.036;
  Matches   74;  Conservative   28;  Mismatches   62;  Indels    0;  Gaps    0;

Qy       1600 ATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCT 1659
Db        736 ATAWAAAATAAAWTAATTWATMAAAAAAAAAAAAAAAAAWATTTRWTTATATWAAWWAAWT 795

Qy       1660 GGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTA 1719
```



```
Db        796 AAWWTWTTKAWWWATWAAATWAAAAAAAAWAAAHWATTAAAAAAAAAAAAAAAAAAWWTAW 855

Qy       1720 ATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTAT 1763
Db        856 WTAWTWAAATYWAAYTTAAAWTTWAWAAAAAATWWWWTTTTTTT 899


RESULT 17
CNS005TE
LOCUS       CNS005TE                997 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL060767
VERSION     AL060767.1  GI:4943573
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 997)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .997
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR12K22"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             2.6%;  Score 51.8;  DB 10;  Length 997;
  Best Local Similarity   18.2%;  Pred. No. 0.084;
  Matches   63;  Conservative  133;  Mismatches  151;  Indels    0;  Gaps    0;

Qy        441 TGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTACTCACTGCTTGTACCTGAGG 500
Db        544 TTYYYYYYYTYYTYYYYTTCYTTYYTYCYYYYCYYTCTCTYTYYYYYYCTYYYCYYTCYTY 603

Qy        501 CCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAGCTGGGTTCCTTATATGGCTC 560
Db        604 YCYYYCYCTYYYYCYTYCYYYYCYYCYYYYYTCYTYYYTTCTCYYTYYYYYTYYAYKYYC 663

Qy        561 TCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTTTTCCTCACCTCTGGGACTGG 620
Db        664 YYYYYTCTCYYYCYCYCYYYTCTCYCYYCYYYYTTCTCTYTTYYTYYYTYYTYTCTCYYY 723

Qy        621 TGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTTTTTCCTGGGGACAATGTCTT 680
Db        724 YYYYYYYTYYTTTCTYYTYYYYTTTYTCTYTYTYTYYTTYYYCYYYCCYCYYYCCYYYY 783

Qy        681 CAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCAGGCTCTTCCTCCTCCAACTG 740
Db        784 CYTCYTYYYYYCTYYYYTYTYTTYTYYCYTYTYTTCTYYYTYTYYYCYCYCYYCTYCCTC 843

Qy        741 GACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAGCATTCTT 787
Db        844 YTYYYCTYCYYYYYCYYYYTCYTYTTMTYTYYYYTYTYTYTYTHYT 890


RESULT 18
CNS016CO
LOCUS       CNS016CO                1200 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106578
VERSION     AL106578.1  GI:5622626
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15E04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             2.6%;  Score 51.8;  DB 10;  Length 1200;
  Best Local Similarity   37.9%;  Pred. No. 0.088;
  Matches  121;  Conservative   51;  Mismatches  145;  Indels    2;  Gaps    1;

Qy       1464 TAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCT 1523
Db        843 TTHGGTTWWWWATAWWAAWAAWAAAAAAAAWATTAAAAATTTTYTAWAAAAATAADGAWTA 902

Qy       1524 CCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATT 1583
Db        903 ATTWTAACGAAAAATTTTAAWAATTAATTWWWAAYAAWAAAAWTWTTTTTWAATTTAATW 962

Qy       1584 TCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA--GGAACAAAAACTACAATCTATG 1641
Db        963 TWTYTTWTAATTTTAAAWWAAAAAWTTTTTTTTWAWATAWAWTAAWAWATAWAAAWTATT 1022

Qy       1642 TAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAA 1701
Db       1023 WAATATAAAAATAWATTWWAWAWAWAWTWWAAWAAAAAAAWAWATWAWAAAAAAWAAAAWW 1082

Qy       1702 AGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTT 1761
Db       1083 WAAAAAAAAAWAWWTTAATAWAWWAWAWTAWAWATATTWTAAAAAWTAWATAWATTWTTT 1142

Qy       1762 ATGATGAGAACTATCTTGA 1780
```

```
                :  : | |  ||| || |
Db       1143 TWWTWWAWATTTATATTTA 1161


RESULT 19
CNS014PQ
LOCUS       CNS014PQ                   987 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL104456
VERSION     AL104456.1  GI:5616067
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN12P22"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             2.5%;  Score 51;  DB 10;  Length 987;
  Best Local Similarity   30.5%;  Pred. No. 0.13;
  Matches   86;  Conservative   68;  Mismatches  128;  Indels    0;  Gaps    0;

Qy       1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499
              |||   |::|: | :|:| :  : :     | ||    :| | ::|||:    |:|| :
Db        687 TAAWTATWWAWNTNWTWTTWWWWTWTTTTTTTTTTTAAWAAAAWWAAAWAAWTWTTAWW 746

Qy       1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559
              :: | | | | |: | ||   |  | |  | | :| : | ||  | : | |  |||
Db        747 ATWWTTTAAAGAAAWTTATTAAATTWAAATTTTWWTTWATTAATTAAWTAATTWATTWA 806

Qy       1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619
              | | ::: | | || : :  |  : | :  |   || | :||   |  :  :::   ||
Db        807 TTWATWWWAATTAATAWTWWTANAKWTTWWTTTTTTTTTWWTATAAATTWTTWWWAWAAA 866

Qy       1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679
              :: :: :: :|  :   :: |::  | :| :: ::    :::::  : || ::  |||
Db        867 TAWRGRWTWWAWAATWAAAWWATWWAAAAWAWWWWDRAAAWDWWWWWAWTAAAWWTTAAA 926

Qy       1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAAT 1721
              || ||||| |||  |: : ::: : ::  | ||| ||:|::
Db        927 AAAAAAAAAWAAAAAAWWWWWWWWWWWTTTTTTTTTTKAWK 968


RESULT 20
CNS0039G/c
LOCUS       CNS0039G                  1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             2.5%;  Score 50;  DB 10;  Length 1101;
  Best Local Similarity   19.4%;  Pred. No. 0.25;
  Matches  129;  Conservative  232;  Mismatches  302;  Indels    3;  Gaps    1;

Qy          4 TTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCA 63
              ::|: |:::::::::::::|   :|:     || ::::: : :|  :  : | :::: :
Db        901 DDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDWWAD 842

Qy         64 GGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAAT 123
              :::  :  ::::  |:| :  ::::: : | :   |:  :: : |::: :||:::::|||
Db        841 DRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDDAAT 782

Qy        124 A---TATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAAC 180
              :   | || : | ::::    :| : | : :::::: :::: | | :: : |::::  :
Db        781 WTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKRR 722

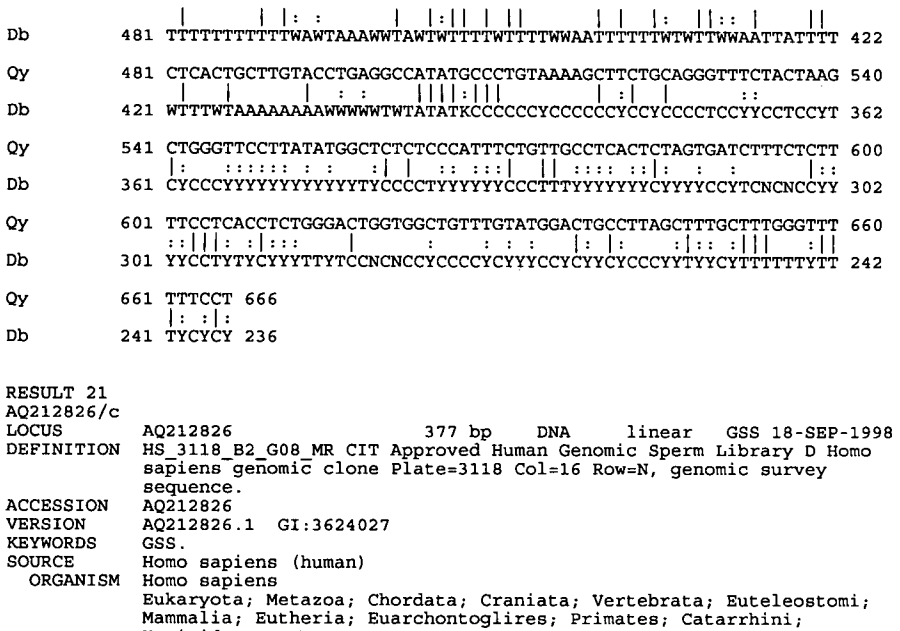
Qy        181 AAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCT 240
              : ::::|: ::|   :::   ::| :|   :  |:: ::   |  :  : :|   :  |
Db        721 DTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAWWTT 662

Qy        241 GTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGA 300
              :  ::::: :: :: |: ::    : | : :|::  :: ::::: : : :::: |: :
Db        661 TDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAADTW 602

Qy        301 TGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTTT 360
              |:|:: :::    :|  |:: :|:   |:::     : |   :|   |:||::  : : |
Db        601 TDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWAT 542

Qy        361 TTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAAC 420
              ||  :| | ||  ||  |||    : |   :  || : | :|    | :  |||||  ||
Db        541 TTATWTTTWTTTWTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAAAAWAAT 482

Qy        421 ATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTA 480
```



```
Db        481 TTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTTTTTWTWTTWWAATTATTTT 422

Qy        481 CTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAG 540

Db        421 WTTTWTAAAAAAAAWWWWWTWTATATKCCCCCCYCCCCCCYCCYCCCCTCCYYCCTCCYT 362

Qy        541 CTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTT 600

Db        361 CYCCCYYYYYYYYYYYYTYCCCCTYYYYYYCCCTTTYYYYYYYCYYYYCCYTCNCNCCYY 302

Qy        601 TTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTT 660

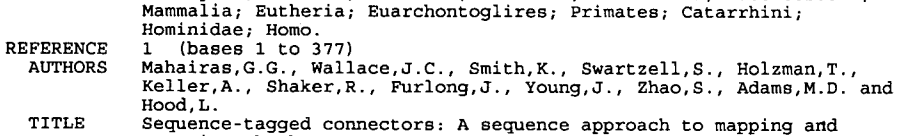
Db        301 YYCCTYTYCYYYTTYTCCNCNCCYCCCCYCYYYCCYCYYCYCCCYYTYYCYTTTTTTYTT 242

Qy        661 TTTCCT 666

Db        241 TYCYCY 236
```

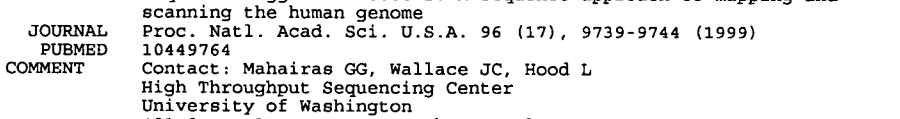
RESULT 21
AQ212826/c
LOCUS AQ212826 377 bp DNA linear GSS 18-SEP-1998
DEFINITION HS_3118_B2_G08_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3118 Col=16 Row=N, genomic survey sequence.
ACCESSION AQ212826
VERSION AQ212826.1 GI:3624027
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 377)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3118 row: N column: 16
Class: BAC ends
High quality sequence stop: 377.
FEATURES Location/Qualifiers
source 1. .377
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3118 Col=16 Row=N"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
ORIGIN

Query Match 2.4%; Score 48.8; DB 9; Length 377;
Best Local Similarity 63.2%; Pred. No. 0.38;
Matches 74; Conservative 0; Mismatches 43; Indels 0; Gaps 0;



```
Qy         55 TATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAA 114

Db        238 TATTCATCATGACTCTCGACTGTACATAAGAAAAACNCAATTCAAGCTGGCTTAAGTATA 179
```

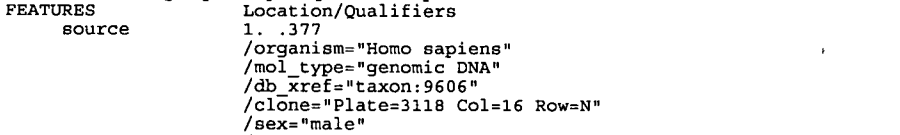


```
Qy        115 AAGGAAAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGC 171

Db        178 AAGGAGAATTGATGAGCCCATCAAATCCAGCCACAGGTAGGGCAAGGAGAGAGTGGC 122
```

RESULT 22
AJ925404
LOCUS AJ925404 1062 bp mRNA linear EST 10-JUN-2005
DEFINITION AJ925404 Theileria annulata merozoite Theileria annulata cDNA clone tam005f04_q1k, mRNA sequence.
ACCESSION AJ925404
VERSION AJ925404.1 GI:67495771
KEYWORDS EST.
SOURCE Theileria annulata
ORGANISM Theileria annulata
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae; Theileria.
REFERENCE 1 (bases 1 to 1062)
AUTHORS Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W., Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M., Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N., Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F., Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F., Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E., Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T., Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J., Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D., Shiels,B., Tait,A., Barrell,B. and Hall,N.
TITLE The genome of the host-cell transforming parasite Theileria annulata and a comparison with T. parva
JOURNAL Unpublished (2005)
COMMENT Contact: Pain A
The Pathogen Sequencing Unit
The Wellcome Trust Sanger Institute
Genome Campus, CB10 1SA, UNITED KINGDOM
Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
FEATURES Location/Qualifiers
source 1. .1062
/organism="Theileria annulata"
/mol_type="mRNA"
/isolate="Ankara (clone D7)"
/db_xref="taxon:5874"
/clone="tam005f04_q1k"
/dev_stage="merozoite"
/lab_host="Bos taurus (cow)"
/clone_lib="Theileria annulata merozoite"
/note="country: Turkey:Ankara"
ORIGIN

Query Match 2.4%; Score 48.8; DB 1; Length 1062;
Best Local Similarity 48.5%; Pred. No. 0.51;
Matches 133; Conservative 0; Mismatches 140; Indels 1; Gaps 1;



```
Qy       1493 TATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATT 1552

Db        765 TATAAATAANNNTAATTTTATNNTTTTTTNACNAGAAATTTTATTTTTATATCTTTTTTT 824

Qy       1553 AGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCAT 1612

Db        825 TATTTTTTAATTATATATTTCATATTTTATTTTTTTGTTTAATTTTNNNCNTTAANACTT 884

Qy       1613 TAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA-GCTGGTTTTGCTCAA 1671

Db        885 TNCCTTATGACTTTAAATTTTAATTTACNTTNTTATCAGTATAATTCTGTTTACACTCAT 944

Qy       1672 TATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTG 1731

Db        945 ATTACTCTTTAAATATAATAATAAATTGTTAATTATAGTNTTATTTTCATCTTGCAGGCT 1004

Qy       1732 AGAATCTATTAAAAAGTATTTGAATTCTTTATGA 1765

Db       1005 ACAATNATTTTATNTGTATTTTATTTTTATATGA 1038
```

```
RESULT 23
CL105409/c
LOCUS       CL105409                1254 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-44M8_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-44M8,
            genomic survey sequence.
ACCESSION   CL105409
VERSION     CL105409.1  GI:40599044
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1254)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 54
            High quality sequence stop: 443.
FEATURES             Location/Qualifiers
     source          1. .1254
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-44M8"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             2.4%;  Score 48.8;  DB 10;  Length 1254;
  Best Local Similarity   60.6%;  Pred. No. 0.54;
  Matches   80;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy       1624 CAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAG 1683
              ||   || | ||| || | ||| |||| || |   ||  ||||   |  ||||||| ||
Db        998 CAGTTACAAAAATATACGGAATAAGAAAAAAATTTTGAATTTGAAAATAATAAAAAGAAA 939

Qy       1684 AAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAA 1743
              | || ||    | | |||   |||||||||  |  |  || | ||||| | | | | |
Db        938 GATAAAAAGAATTTTTAAATGTAAAATATTTAATATAAAAGATAATTGATATTATTTGAG 879

Qy       1744 AAAGTATTTGAA 1755
              ||  |||| |||
Db        878 AAGTTATTAGAA 867


RESULT 24
AQ338558/c
LOCUS       AQ338558                391 bp    DNA     linear   GSS 12-JAN-1999
DEFINITION  HS_3118_B1_G08_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3118 Col=15 Row=N, genomic survey
            sequence.
ACCESSION   AQ338558
VERSION     AQ338558.1  GI:4144574
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 391)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  PUBMED    10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3118  row: N  column: 15
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 391.
FEATURES             Location/Qualifiers
     source          1. .391
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3118 Col=15 Row=N"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

  Query Match             2.4%;  Score 48.2;  DB 9;  Length 391;
  Best Local Similarity   63.2%;  Pred. No. 0.55;
  Matches   74;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy         55 TATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAA 114
              |||   ||| |||  | || || |  | | | |||| ||||||||  |   | |||||  |
Db        238 TATTCATCATGACTCTCGACTGTACATAAGAAAAACCCAATTCAAGCTGGCTTAAGTATA 179

Qy        115 AAGGAAAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGC 171
              ||||| |||  ||  || |||  || |    |||||  |||||| |  |  || |||
Db        178 AAGGAGAATTGATGAGCCCATCAAATCCAGCCACAGGTAGGGCAAGGAGAGAGTGGC 122


RESULT 25
CNS016E2/c
LOCUS       CNS016E2                1204 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106628
VERSION     AL106628.1  GI:5622852
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1204)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
```

```
     source          1. .1204
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15A12"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             2.4%;  Score 48.2;  DB 10;  Length 1204;
  Best Local Similarity   20.3%;  Pred. No. 0.76;
  Matches   70;  Conservative  134;  Mismatches  141;  Indels    0;  Gaps    0;

Qy       1593 ACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGA 1652
Db       1175 AWWKAKAKWKAAAAAAAAAAAAAAAAAAKAAAADAAAAKKKAAAAAAAAKAADDAAADAAA 1116

Qy       1653 ACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATAT 1712
Db       1115 AAKAKAAKAKAAAAAAAADAAWKKKAWAAAAAAKWAAAKWKATKKDDDDKDKAAADAKKKK 1056

Qy       1713 TTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAAC 1772
Db       1055 KKAKADADKAADKDKDKKAKADAWDDADDAAKKKKKAKKKDKKKDDATDADWAAKAAKKK 996

Qy       1773 TATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGA 1832
Db        995 DKKAWKKKKDKKKAKKAKAWKKKDKKDKKDKKKKKKKDKKAKKKKKAAAKDAAAAAAAAA 936

Qy       1833 GGCTTTGTCATAAAGTCTTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTT 1892
Db        935 KKKDKKKDAAKAAAKKKKKDAAKKKKAKKKDDAWAADADKAAAKKKAAAKKKDKKKDAAD 876

Qy       1893 AAAAGCATACTGGATTTTGTTTTAGACTTTTAGTGAACTGAAGGG 1937
Db        875 KKAAKDAAKKKKDKAAKAADKDAAKKWDKKKAGGAARAAAKKGK 831


RESULT 26
CW951690
LOCUS       CW951690                 958 bp    DNA     linear   GSS 20-DEC-2004
DEFINITION  TcB37.2_A08_SP6 Tribolium BAC library Tribolium castaneum genomic,
            genomic survey sequence.
ACCESSION   CW951690
VERSION     CW951690.1  GI:56734727
KEYWORDS    GSS.
SOURCE      Tribolium castaneum (red flour beetle)
  ORGANISM  Tribolium castaneum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Tenebrionidae; Tribolium.
REFERENCE   1  (bases 1 to 958)
  AUTHORS   Savard,J. and Tautz,D.
  TITLE     Tribolium castaneum BAC-ends sequencing project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Savard, J.
            Abteilung fur Evolutionsgenetik, AG Tautz
            Institut fur Genetik, Universitat zu Koln
            Weyertal 121, 50931 Koln, Germany
            Tel: 49 221 470 6911
            Fax: 49 221 470 5975
            Email: savard@uni-koeln.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .958
                     /organism="Tribolium castaneum"
                     /mol_type="genomic DNA"
                     /strain="GA-2"
                     /db_xref="taxon:7070"
                     /clone_lib="Tribolium BAC library"
                     /note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1;
                     Library constructed by Exelixis Inc."
ORIGIN

  Query Match             2.4%;  Score 48;  DB 10;  Length 958;
  Best Local Similarity   50.4%;  Pred. No. 0.8;
  Matches  117;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       1540 TAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATC 1599
Db        517 TAAAAAAATTAAAAAAGAGATATAAAAATAATGATATATTTGATAATAGGATAAAATTAA 576

Qy       1600 ATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCT 1659
Db        577 AAAAAATTTAAATAAAATAATAATATAATTAATAGATAATAATATTATTAGAAAAATAT 636

Qy       1660 GGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTA 1719
Db        637 AAATAATCAAATAAAATATATAAAAAAAAAAAAAAAAATAATAATAAAAATATTTAATAA 696

Qy       1720 ATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAA 1771
Db        697 AAAAGGTAAAATAAATAAATTTATAATGAATTATAAAAATTAAAAATTAAAA 748


RESULT 27
CNS020K7
LOCUS       CNS020K7                 1092 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1
                     end : T7"
```

```
ORIGIN

  Query Match             2.4%;  Score 48;  DB 10;  Length 1092;
  Best Local Similarity   37.9%;  Pred. No. 0.83;
  Matches   72;  Conservative   39;  Mismatches   79;  Indels    0;  Gaps    0;

Qy       1596 TATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACA 1655
              ::: | ||  ||   | ::   |     | |||||  : |:: ::: :    | || ||
Db        627 WWWWAAAAAAAAAAAAAAWWTTTTTTTTAWAAAAAAAWAAWWWWWWWWTTAAAAAAAAAT 686

Qy       1656 AGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTG 1715
                |  ||||  |   | ||||||  || |||||  ||:  |  : ||:   ||||:| :
Db        687 TTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAWTWAATWAAAWWAAAAAAWAAWAA 746

Qy       1716 TTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTAT 1775
              :::|:| :   |  |: : ::: :|  |:||| ::|:| |:|| | :::  : :    |||
Db        747 WWWAWTTWTTTTTTWAWWWWWWWAAAAWAAAWWWTWTTAWTTTTAWWWATWAWWTTWTAT 806

Qy       1776 CTTGACTCAA 1785
                || | : ||
Db        807 TTTAAAWAAA 816


RESULT 28
AL551397
LOCUS       AL551397                950 bp    mRNA    linear   EST 30-MAR-2004
DEFINITION  AL551397 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI064YC04 5-PRIME, mRNA sequence.
ACCESSION   AL551397
VERSION     AL551397.3  GI:45856202
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:31273213.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4987.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DI064BB02QP1&c=4987.r.
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI064YC04"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             2.4%;  Score 47.8;  DB 1;  Length 950;
  Best Local Similarity   36.5%;  Pred. No. 0.9;
  Matches   58;  Conservative   42;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
                | : | : :||   : ::   :|  ||:|:||:::| |:  | :|| :  ::|||:|:|

Db        259 AAMTTCMARAAAGGMAWWAAAMAMMAAMARAAWMWAAAMAATTWGTTWMAWRAAGMAMAT 318

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
                | |    ||    |: : |:||| || :::||:::::||  ||:|:   |||:| | | |
Db        319 GGTTWCCTTMAAAAMARAAMAAAAAAMWWWAARMMWMCAMATGWARAAAAAAWTGTGTTT 378

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAA 1755
                |:|   |   |   |   |:|:  ||:|:| |:::  |:
Db        379 GTRAAATGAARAAMWATTGTMTWAAAAWAWGGAWWWWAW 417


RESULT 29
CL071552
LOCUS       CL071552                1140 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-121C5_RM1.1 CH216 Xenopus tropicalis genomic clone
            CH216-121C5, genomic survey sequence.
ACCESSION   CL071552
VERSION     CL071552.1  GI:40527465
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1140)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 43
            High quality sequence stop: 164.
FEATURES             Location/Qualifiers
     source          1. .1140
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-121C5"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             2.4%;  Score 47.6;  DB 10;  Length 1140;
  Best Local Similarity   47.6%;  Pred. No. 1.1;
  Matches  140;  Conservative    0;  Mismatches  154;  Indels    0;  Gaps    0;

Qy       1478 TACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATA 1537
                ||   || |||     ||  |  |   | | | | || | | |  |  | | |   ||
Db        468 TAATAATTAAATAAAAATAAATTATAATTAAAAAAAATTATATATTATATATATAATATT 527

Qy       1538 TGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTA 1597
                |||| |   |  | | |   | ||||  | | ||   |  | ||   |  | | |
Db        528 AATAAAAATAATAAAAAAATAATTTAAATCAAATAATTAATATAATTACATAATAAAAAA 587

Qy       1598 TCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAG 1657
                |||  || | | || || |  ||  | |  |  |||  || ||||||| || || ||
Db        588 AATTAAATAATTTAATATCTTATTAAACATATTTTAAATAAATATAATTAAAAAAAAAAA 647

Qy       1658 CTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTT 1717
                |       | ||| ||| ||  |||| |||  |   |||   |||||| || ||
Db        648 AATAAATAATAATAGATATAAAAAAATAAAATAAATAAAAATAAAAATAAAATAATTATT 707

Qy       1718 TAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAA 1771
```



```
Db        708 ATCATAATATATATAAAAAAATAACAACTATATTATAAATTTTTAATATATTAA 761


RESULT 30
AU062068/c
LOCUS       AU062068               548 bp    mRNA    linear   EST 20-MAY-1999
DEFINITION  AU062068 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLH458, mRNA sequence.
ACCESSION   AU062068
VERSION     AU062068.1  GI:4883172
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLH458"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

  Query Match             2.4%;  Score 47.4;  DB 1;  Length 548;
  Best Local Similarity   54.2%;  Pred. No. 0.98;
  Matches   96;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy       1606 AAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTT 1665
Db        249 ATGAAATTAAAAAAAAAAAAAAAAAAAAGTGTTAATAAATTGAGAACATTAAAGAAGTTTT 190

Qy       1666 GCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGG 1725
Db        189 TTTTTATATATAATGAAAAAAAAATAAATAAATGAAAAAATAAAAAATTAAATAATAAAA 130

Qy       1726 TCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACT 1782
Db        129 AGAGTTTATTTATAATAAACACTCATATTATTTTTTATTTCAAATTAATTTTTTCT 73


RESULT 31
CNS015YI/c
LOCUS       CNS015YI               1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15K02 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106068
VERSION     AL106068.1  GI:5619869
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15K02"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             2.4%;  Score 47.4;  DB 10;  Length 1201;
  Best Local Similarity   31.7%;  Pred. No. 1.2;
  Matches  137;  Conservative   89;  Mismatches  203;  Indels    3;  Gaps    1;

Qy       1362 TAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTAT 1421
Db       1200 TMTMYTATCTTMTCTTCHYWYTYTCTCWCATCWCTCTTHTYTHYTCTCACAYAMHTTTTY 1141

Qy       1422 TATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACA 1481
Db       1140 TAYCHATTHTCYWHMYTTYTTTAWTMTATATAYAATHTTCWHAAWAWAACATATMTWMTA 1081

Qy       1482 TATAAAA---GGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATAT 1538
Db       1080 KATCAAHHWYMMMMMTTAYYCYAMYAYTTAWTTTWTTTTTTAHMMCYAAWWWCACMYTC 1021

Qy       1539 GTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTAT 1598
Db       1020 YAWMTWAAYAATMMHMWMWHAAMTATAWTTACMTCTTTWHHAATCCWTATMAACCMACAW 961

Qy       1599 CATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGC 1658
Db        960 AAWAAAMAAWWMACTHMMMMMMMAAAAAMMACCCCCCCYYYYTMAAAAMACCWAAAAAAA 901

Qy       1659 TGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTT 1718
Db        900 AAAAAAAAWCCAATAAACTAACCATAAAACAWAMAAAAAMAAAAMWHAAAKBHYYYTTTT 841

Qy       1719 AATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTT 1778
Db        840 YYAAWTATWWTYCKCARAAMATGNBSMTTTAATCRACYCCYTCATYHWTTAATCACACBT 781

Qy       1779 GACTCAAGTGGA 1790
Db        780 CABTTACBKGDA 769


RESULT 32
CNS0161D
LOCUS       CNS0161D               1225 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106171
VERSION     AL106171.1  GI:5620504
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1225)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1225
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15C18"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             2.4%;  Score 47.2;  DB 10;  Length 1225;
  Best Local Similarity   27.7%;  Pred. No. 1.4;
  Matches  130;  Conservative  101;  Mismatches  239;  Indels    0;  Gaps    0;

Qy       1286 TCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAAT 1345
Db        551 KNNKNKKTKCTCKNNKCATCATTTTWTWATTTTWTTWWTTTTTTTATTTTTWTYATYAAT 610

Qy       1346 TTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTT 1405
Db        611 AAWTWTTTTWTTTAAYYMCCKYTTWTAAAAAAATTAACCTCCMYYTHYYAWTAYMWTCNT 670

Qy       1406 GCTAATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTA 1465
Db        671 TMTWHYYTTCCYTCCYYCCCTCCCCCTCTCYTTYTYTTTTYTHTCTTTTTTTTTYTHAHAW 730

Qy       1466 AGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCC 1525
Db        731 CCKYYYCHMMAAMAMAYAKCCKBKKKNNKWCKMKMKHMAADAAKAMAMNWAAHHWTTTYT 790

Qy       1526 ACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTC 1585
Db        791 WAAHAYNAAAAATTTAAAATCKCAAAMAMAAMAAAMMAAAAACCCMCCVCCKKKKDKKKK 850

Qy       1586 TCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAAT 1645
Db        851 KKKKDDKDAAKRARADDRKAAAAAAAAATATAAAAAAAAAAAAAAAAAWWWATTYTYYYAY 910

Qy       1646 TAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTC 1705
Db        911 AWAAAYAWATAAWWAWTTWTTAAAWWAAAAAAAAAWAWWTTTYAAAAAAWTWTWAAATAA 970

Qy       1706 AAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAA 1755
Db        971 WTTATTTTTTTTTAWWWTTTATWWATAATTTATWTTTWTTTTTTTTWTAA 1020


RESULT 33
CL022041
LOCUS       CL022041                1281 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-13E20_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-13E20, genomic survey sequence.
ACCESSION   CL022041
VERSION     CL022041.1  GI:40463854
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1281)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 142
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1. .1281
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-13E20"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             2.3%;  Score 46.8;  DB 10;  Length 1281;
  Best Local Similarity   52.0%;  Pred. No. 1.8;
  Matches  129;  Conservative    0;  Mismatches  117;  Indels    2;  Gaps    1;

Qy       1501 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 1560
Db        497 AATTTATTAAAAAAGGATTTTACAAATAAGTATAATGTGTCAACATATAAATTAAT--AT 554

Qy       1561 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 1620
Db        555 AGAAAGTAATTAATAATTCTAATTATAAAGAAATTCATTTTTCATAACTAATCAACTATA 614

Qy       1621 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 1680
Db        615 AATGAAAAACAATTATTTATGTAATAAATTTCACATATTATTACAATCATAGAAAAAGAT 674

Qy       1681 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 1740
Db        675 AAAAATAACCAATTATATATAAATTATAATATTTTATTAAAAATATAAAAGCTAATAATT 734

Qy       1741 TAAAAAGT 1748
Db        735 AAAGTAAT 742


RESULT 34
DR071573
LOCUS       DR071573                511 bp    mRNA    linear   EST 08-JUN-2005
DEFINITION  RTDK1_20_F06.g1_A029 Roots, dark Pinus taeda cDNA clone
            RTDK1_20_F06_A029 5', mRNA sequence.
ACCESSION   DR071573
VERSION     DR071573.1  GI:67049321
KEYWORDS    EST.
SOURCE      Pinus taeda (loblolly pine)
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Pratt,L., Cordonnier-Pratt,M.M., Lorenz,W.W., Zimmermann,C. and
            Dean,J.F.D.
  TITLE     An EST database from dark-treated loblolly pine (Pinus taeda) roots
  JOURNAL   Unpublished (2005)
```

```
COMMENT     Other_ESTs: RTDK1_20_F06.b1_A029
            Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            RNA prepared and library constructed by W. Walter Lorenz (School of
            Forest Resources, University of Georgia); plant material prepared
            by Craig Zimmermann (School of Forest Resources, University of
            Georgia) using rooted cuttings provided by the Forest Biology
            Research Cooperative (FBRC) and the CCLONES project a the
            University of Florida; sequencing done in the Laboratory for
            Genomics and Bioinformatics, University of Georgia. Sequence ends
            have been trimmed to exclude vector and regions below Phred quality
            16. Three-prime sequences are presented as their reverse complement
            and have been trimmed to exclude polyA.
            Seq primer: JENREV (CAGGAAACAGCTATGACC).
FEATURES              Location/Qualifiers
     source           1. .511
                      /organism="Pinus taeda"
                      /mol_type="mRNA"
                      /strain="3 CCLONES"
                      /db_xref="taxon:3352"
                      /clone="RTDK1_20_F06_A029"
                      /lab_host="DH10B-T1 phage-resistant E. coli"
                      /clone_lib="Roots, dark"
                      /note="Organ: Root; Vector: pSL1180; Site_1: EcoRI;
                      Site_2: XhoI; The library was prepared from polyA+ RNA
                      from the roots of 1-year-old loblolly pine (Pinus taeda)
                      cuttings that were rooted and then planted in washed sand.
                      The rooted cuttings were maintained for 50 days (May 1
                      2003 harvest) under ambient conditions in a local
                      greenhouse. They were kept on a weekly regimen of 0.5x
                      nutrient-complete Hoagland's solution and supplemented
                      with additional water sufficient to maintain a 15% soil
                      moisture content. Twenty-four hours (24h) prior to
                      harvesting roots for mRNA preparation, the potted trees
                      were placed in a dark growth chamber at 25 C.
                      Double-stranded cDNA was cloned unidirectionally into
                      pSL1180. Inserts can be excised with EcoRI (5' end) and
                      XhoI (3' end)."
ORIGIN

  Query Match             2.3%;  Score 46.6;  DB 8;  Length 511;
  Best Local Similarity   57.5%;  Pred. No. 1.5;
  Matches  103;  Conservative    0;  Mismatches   74;  Indels    2;  Gaps    1;

Qy       1596 TATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACA 1655
Db          2 TACCAAGTAGAATTTCATGGATTAGGACCAATTACAATACACTTTGTAATTTTAAAACAA 61

Qy       1656 AGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTG 1715
Db         62 TACCATTTAACATTAAAATATAATTAACCCAAAAAAACCATTACCATGAAAAACAATTAA 121

Qy       1716 T--TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAAC 1772
Db        122 TAATTAATCACCACAAAGAGAATCCAGTCCCAAGTAATCATTTCATGGATTATGTGAAC 180


RESULT 35
CV569142/c
LOCUS       CV569142               617 bp    mRNA    linear   EST 22-OCT-2004
DEFINITION  od01h09.y1 Human keratoconus cornea, unamplified, (od/oe) Homo
            sapiens cDNA clone od01h09 5', mRNA sequence.
ACCESSION   CV569142
VERSION     CV569142.1  GI:54468336
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Rabinowitz,Y., Dong,L. and Wistow,G.
  TITLE     Expressed sequence tag analysis of human keratoconus cornea
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 01  row: h  column: 09
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES              Location/Qualifiers
     source           1. .617
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="od01h09"
                      /tissue_type="Cornea"
                      /dev_stage="Adult"
                      /lab_host="EMDH10B"
                      /clone_lib="Human keratoconus cornea, unamplified,
                      (od/oe)"
                      /note="Organ: Eye; Vector: pCMVSport6; Approximately 40ug
                      total RNA was extracted from 7 adult human keratoconus
                      corneas. A directionally cloned cDNA library in the
                      pSPORT1 vector (Invitrogen) was constructed at Bioserve
                      Biotechnology (Laurel MD) essentially following the
                      protocols of the SuperScript Plasmid System full details
                      of which are contained in the manufacturer's Instruction
                      manual (http://www.lifetech.com/). First strand synthesis
                      was carried out using a Not I primer-adapter
                      [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. cDNA was
                      cloned in Not I/Sal I sites. EST analysis was performed on
                      the unamplified library at the NIH Intramural Sequencing
                      Center (NISC)."
ORIGIN

  Query Match             2.3%;  Score 46.4;  DB 7;  Length 617;
  Best Local Similarity   49.7%;  Pred. No. 1.8;
  Matches  171;  Conservative    0;  Mismatches  171;  Indels    2;  Gaps    2;

Qy       1448 AATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGA 1507
Db        617 AATGCTTTGTTTTCTAAAACTTGTATTGTTTTTAAACAAAAATAGATTTGAAAAGAATGA 558

Qy       1508 TACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATA 1567
Db        557 ACCACAACAGTGCTACCAAAAAGTCACATCTTGTTACCAGACATCA-CTGTCCTTACAAC 499

Qy       1568 GACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAA 1627
Db        498 AATTCAACTAATAAGATTTAAGATTTAACTTCATCCTAATGTATCAGAACATTAAAAAAA 439

Qy       1628 AACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAA 1687
Db        438 AAAAAGTCTCAATGTACCTATTTAAATAAATCGTGCTTCCATAATAATATTCTTTAAAAA 379

Qy       1688 AGAAACCATGTGAA-AGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAA 1746
Db        378 TGAAAAACAGTAAAGAGCACACATTTTTATTTACTCACAACACTGAATAATTAATGTAAA 319

Qy       1747 GTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGA 1790
Db        318 CTTTTTGAATTTTTTTTTTCTTTAGACATTTTTCCTCTAGAGTA 275


RESULT 36
BH712636
LOCUS       BH712636               644 bp    DNA     linear   GSS 20-FEB-2002
```

```
DEFINITION  BOMFC43TF BO_2_3_KB Brassica oleracea genomic clone BOMFC43,
            genomic survey sequence.
ACCESSION   BH712636
VERSION     BH712636.1  GI:18804083
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 644)
  AUTHORS   Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
            Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea and its
            application to gene discovery and annotation in Arabidopsis
  JOURNAL   Genome Res. 15 (4), 487-495 (2005)
   PUBMED   15805490
COMMENT     Other_GSSs: BOMFC43TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMFC43"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             2.3%;  Score 46.4;  DB 9;  Length 644;
  Best Local Similarity   53.3%;  Pred. No. 1.9;
  Matches   98;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
              | |  | ||||    |  |    |  ||| ||||| |||||  |   ||| || || || |
Db        451 ACCGAAGGGAAAATACAAAGGAGTGAAAAAAAATTACAAAATTCATAAATAAAATAAGAG 510

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
                ||||  | |||  | | | |  |||||  | |  || ||  | |   || || |
Db        511 AAACATTTTATTTAATTCATAGAAAGCAAAAACACATTATATGTTATTTGGAAGATGTCC 570

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATC 1776
                | | |  ||||  |   || || |||   | ||| |||| || |||   |     |||
Db        571 AAATTTATCTCATATATGTTCAATCAAATCTTCTTTTAATTGTTGATGGATATGTGGATC 630

Qy       1777 TTGA 1780
                ||
Db        631 CCGA 634


RESULT 37
AZ899171
LOCUS       AZ899171                649 bp    DNA     linear   GSS 05-MAR-2001
DEFINITION  RPCI-24-213I16.TJ RPCI-24 Mus musculus genomic clone
            RPCI-24-213I16, genomic survey sequence.
ACCESSION   AZ899171
VERSION     AZ899171.1  GI:13218116
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 213  row: I  column: 16
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-213I16"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
ORIGIN

  Query Match             2.3%;  Score 46.4;  DB 9;  Length 649;
  Best Local Similarity   52.5%;  Pred. No. 1.9;
  Matches  124;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy       1595 TTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAAT-TAGAAGAA 1653
              | | ||    ||||   |||   ||| |||| |||  ||||||  |     |  |||| ||
Db        326 TAAACAGCTAGAAGCACTTAAAGAAGAAACACAAAACACAATCAAACAGGTGAAGAAAAA 385

Qy       1654 CAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATT 1713
                ||        | |   || |||||||  | |||||||||| |||   || ||| || |
Db        386 GAACAAAACCATCCAATATCTAAAAATGAAAAAAAGAAACAATGAAGAAATCACAAAAGA 445

Qy       1714 TGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACT 1773
                | | |  | |||  | || || |||  ||| ||     || |  |  |||  || | |
Db        446 AGATAACCCTGGTGTTGGAAAACCTAAGAAAGAGATCAGGAGTCATAGATGCAGACATCA 505

Qy       1774 ATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAA 1829
                | |  |    ||||  | |||  | || |     |       | ||||||
Db        506 CCAAAAAATACAAGATACAGAAGAGAGAATCTTAGGTGCAGAAGATACCATAGAAA 561


RESULT 38
AZ899889
LOCUS       AZ899889                733 bp    DNA     linear   GSS 05-MAR-2001
DEFINITION  RPCI-24-213J16.TJ RPCI-24 Mus musculus genomic clone
            RPCI-24-213J16, genomic survey sequence.
ACCESSION   AZ899889
VERSION     AZ899889.1  GI:13218747
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 733)
```

```
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-213J16.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 213  row: J  column: 16
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .733
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-213J16"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
ORIGIN

  Query Match             2.3%;  Score 46.4;  DB 9;  Length 733;
  Best Local Similarity   52.5%;  Pred. No. 1.9;
  Matches  124;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy       1595 TTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAAT-TAGAAGAA 1653
Db        365 TAAACAGCTAGAAGCACTTAAAGAAGAAACACAAAACACAATCAAACAGGTGAAGAAAAA 424

Qy       1654 CAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATT 1713
Db        425 GAACAAAACCATCCAATATCTAAAAATGAAAAAAAGAAACAATGAAGAAATCACAAAAGA 484

Qy       1714 TGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACT 1773
Db        485 AGATAACCCTGGTGTTGGAAAACCTAAGAAAGAGATCAGGAGTCATAGATGCAGACATCA 544

Qy       1774 ATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAA 1829
Db        545 CCAAAAAATACAAGATACAGAAGAGAGAATCTTAGGTGCAGAAGATACCATAGAAA 600


RESULT 39
CNS0155H
LOCUS       CNS0155H                1001 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN13C23 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL105023
VERSION     AL105023.1  GI:5617037
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1001)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1001
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN13C23"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             2.3%;  Score 46.4;  DB 10;  Length 1001;
  Best Local Similarity   38.7%;  Pred. No. 2.1;
  Matches   99;  Conservative   46;  Mismatches  106;  Indels    5;  Gaps    1;

Qy       1535 ATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCAC 1594
Db        720 ATAARWAARARDARAAAAAAATTAWWTAAWWWAWAAWWWWAAATAAAAADTTWTATWAAA 779

Qy       1595 TTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAAC 1654
Db        780 TAAAAAAAWWTAAAAAAATATTTAWAWAWAAAAATWAAAAATTWWAAAAWWTAWWAAAAA 839

Qy       1655 AAGCTGGTTTTGCTCAATATAA-----AAATAAGAAAAAGAAACCATGTGAAAGTCAAAA 1709
Db        840 AAATTWAATTTTTTAAAWAWAAWWAATWAWATATAAAADAAAAAAAAAAWAAWATAAAAT 899

Qy       1710 TATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAG 1769
Db        900 TWAAAAWTTWWWTAWAWADWTWTTTWTTWTWWAAWAAATAAWTWTAWTTAATWTAAWAAT 959

Qy       1770 AACTATCTTGACTCAA 1785
Db        960 AAAWATAAAAAATWAW 975


RESULT 40
CG757986
LOCUS       CG757986                1364 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P053-2-D02.zb Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG757986
VERSION     CG757986.1  GI:37987099
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1364)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
```

```
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1364
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

  Query Match             2.3%;  Score 46.2;  DB 10;  Length 1364;
  Best Local Similarity   43.5%;  Pred. No. 2.6;
  Matches  217;  Conservative    0;  Mismatches  279;  Indels    3;  Gaps    1;

Qy       1285 ATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAA 1344
              ||  |  ||| | |  |       ||||||  |  ||||   |||  |||| | | ||||
Db        545 ATTNTTATTTATTATTATATATTAAATTTTTTTTTATAAAANTTTATAATTTAATAATAA 604

Qy       1345 TTTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTT 1404
              || |   | | |    |||   |  || |   ||||        |  || |
Db        605 ATTATTNTAATATTATTTAAAATAAATANANTAAATATTTNAAATTTTAAAATAATTANN 664

Qy       1405 TGCTAATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACT 1464
                | ||| || | ||| |     |  | |  || |  |    || | ||  || | |
Db        665 NTTTTATAAATTTTTATAAAATAATATTTTTTTTTTNTATTAAAATTATTAAATTTTAAAA 724

Qy       1465 AAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTC 1524
                |  ||   ||| ||| ||||    |  | |  | |    || |   | |  | |
Db        725 TTTTTTTTAATTTAAATAAAAAATAATTTATAAATATTAAATTAAAATTATTTAATATT 784

Qy       1525 CA---CAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCA 1581
               |    |  | | ||  |  ||    ||  ||  |||   | || || | |    |   |
Db        785 TAATTTAANTATAATTTAAAATATTTTTTTTTAATTTAATTAATTAAATTATATTNNTTA 844

Qy       1582 TTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATG 1641
              |||  |    | |  |||  |    |  |    |  ||  |    |  |  | | |||
Db        845 TTTATTATTTTATATATTTTTTTAATATTTAAAAAAAAATATTTTTATNTTTATTATATA 904

Qy       1642 TAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAA 1701
              || | |          |  ||||    ||  |  |    |  |           | | ||
Db        905 TATTAAATTATTTTTTTTAATTTTAAAATATTTTTATTATATTATTTTTTTTTNTTTTAA 964

Qy       1702 AGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTT 1761
              |   ||| |||| |||  |    | ||| | | | || |||||  || ||   || ||
Db        965 ATATTAAAAATTTTTTTTTTTTTATTTATTAATATTATAATAAAATTTAATTTTTTTTTN 1024

Qy       1762 ATGATGAGAACTATCTTGA 1780
              |  | |  | ||| || |
Db       1025 TTTTTTATTATTATATTAA 1043


RESULT 41
CV784532/c
LOCUS       CV784532                778 bp    mRNA    linear   EST 10-NOV-2004
DEFINITION  UI-M-HY0p-ctb-f-07-0-UI.r1 NIH_BMAP_HY0p Mus musculus cDNA clone
            IMAGE:30696222 5', mRNA sequence.
ACCESSION   CV784532
VERSION     CV784532.1  GI:55689472
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 778)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .778
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30696222"
                     /tissue_type="whole eye"
                     /dev_stage="embryos 15,16,17,18 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_HY0p"
                     /note="Organ: Eye; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site .Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequences located between the Not I site and the polyA
                     tail are CTGCGTCCTC and AGAGGAGACG. This library was
                     created for the University Iowa Brain Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Institute of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
ORIGIN

  Query Match             2.3%;  Score 46;  DB 8;  Length 778;
  Best Local Similarity   49.2%;  Pred. No. 2.5;
  Matches  118;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy       1684 AAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAA 1743
              ||||||||       |  ||  |  ||| ||||    |||      |  | ||    | |
Db        777 AAAAAGAACAGGACTAGAAAACTCAATTTTTGACATATCTAAAAAGGGGGACAAGAACAG 718

Qy       1744 AAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTT 1803
              |  |||||   ||    ||  ||  |   ||   ||    | | ||  || || |
Db        717 TAGTCATTTGTTCTCNCCGTGGNGACAGGAATGAAGAGGAAGGAGGGAAGAGGGGCTTCA 658

Qy       1804 TTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTACAGG 1863
              |||   ||||    |   |     |   |||||   |     | ||| ||  | || |  ||
Db        657 TTTCTTCTGTTCACAAGAGAATGCTTGGAGGGGATTCTTCAGAGTTTTTCAGGGGATGGG 598

Qy       1864 TGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTT 1923
                ||||  |  ||||||||  ||  |  ||  ||      |  | |||||||| ||  |  | |
Db        597 ACCCAAATGTAGTGCCCAGACTGGTCCTCAATGATTTGTTTGATTTTGTTATAAATCTCT 538


RESULT 42
DN573331/c
LOCUS       DN573331                889 bp    mRNA    linear   EST 14-MAR-2005
DEFINITION  92944934 Sea Urchin primary mesenchyme cell cDNA library
            Strongylocentrotus purpuratus cDNA clone PMCSPR2-77O21 3', mRNA
            sequence.
ACCESSION   DN573331
VERSION     DN573331.1  GI:61132370
KEYWORDS    EST.
SOURCE      Strongylocentrotus purpuratus
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
```

```
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
            Ettensohn,C.A.
  TITLE     A large scale analysis of mRNAs expressed by primary mesenchyme
            cells of the sea urchin embryo
  JOURNAL   Development 128 (13), 2615-2627 (2001)
   PUBMED   11493577
COMMENT     Contact: Erica Sodergren
            Human Genome Sequencing Center
            Baylor College of Medicine
            One Baylor Plaza, Houston, TX 77030, USA
            Tel: 713-798-7676
            Fax: 713-798-6977
            Email: ericas@bcm.tmc.edu
            NCBI Trace Archive: 496943148
            Insert Length: 1750  Std Error: 0.25
            Plate: 77  row: O  column: 21.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Strongylocentrotus purpuratus"
                     /mol_type="mRNA"
                     /db_xref="taxon:7668"
                     /clone="PMCSPR2-77021"
                     /tissue_type="embryo"
                     /cell_type="primary mesenchyme cells"
                     /lab_host="E.coli"
                     /clone_lib="Sea Urchin primary mesenchyme cell cDNA
                     library"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: MluI; oligo
                     dt priming from poly A+ RNA, directionally cloned"
ORIGIN

  Query Match             2.3%;  Score 46;  DB 8;  Length 889;
  Best Local Similarity   50.0%;  Pred. No. 2.6;
  Matches  115;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       1482 TATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTA 1541
              | | ||| |   |   |  |||   | | |  |  |   || || | | | |  ||| ||
Db        246 TTTTAAATGCCCAAAAAAAAATCCCAAAAACCAAACCCCTCCCCCCCGTTTTTTTATTTA 187

Qy       1542 AAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCAT 1601
              || || |  ||  ||     ||| | |   |  | |  |||| ||  | ||   |  |
Db        186 AAAAACTTTTTTTTTTTCAAAAAAAAAAAAAAAATTTAAATTTTCCCCCCAGAAAAAAA 127

Qy       1602 AAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGG 1661
              ||  ||   |  |   ||  || |||||  | ||   |   ||  | || || ||
Db        126 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAC 67

Qy       1662 TTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATA 1711
              ||      || | ||||| || ||||| |||   |    ||| | |||| |
Db          66 TTAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAAAAAGA 17


RESULT 43
CNS00DKY/c
LOCUS       CNS00DKY                928 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidae; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR27A24"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             2.3%;  Score 46;  DB 10;  Length 928;
  Best Local Similarity   30.2%;  Pred. No. 2.6;
  Matches  110;  Conservative   74;  Mismatches  180;  Indels    0;  Gaps    0;

Qy       1409 AATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGG 1468
              :: ::: : : :||:: : : :: : ::| ||      |   | |||| |::  : :::
Db        838 MMMMMMMMMMMTTMWWHHHWHHHHHHWTTTTTTWTTTTTTTTTTTTWWWWWWWWWW 779

Qy       1469 CTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACA 1528
              | ::   ::: ::::::        :: :  ::: : |||   :  | |: : |  | |
Db        778 TTTWWWWTWWWWWWWWWTNNNWNMCHWAWATWWWWWATATTTWTATWTWAWATAWAAA 719

Qy       1529 GGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCT 1588
                | |  : ||   || :||  |:|     || |       : | |  :||| : |
Db        718 AAAAAAAAAAAWTAAATTTTTWTTTAWTTWTAAAATTTTTTATTWTTTTTTWTTTTWTT 659

Qy       1589 GCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAG 1648
                |  | |: :||:  ||   ||    ||  :| |   : ||  :  |||  |||::
Db        658 TWTATATAAWAWTAWAWAAWAWATATTATATWAWATATTTWWTATTWAWTATAAAATWWT 599

Qy       1649 AAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAA 1708
              |:    ||      | :  | || |: :|||:|| ||||| |||  |: | : :   ||:
Db        598 AWWTTTAAWTAAWATAWAATTAAAAWTWAAAWAAAAAAAAAWAAAATAWTTTWTWAAAAAW 539

Qy       1709 ATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGA 1768
              : |     ||: |    :     | ||  | | : |:  |||||  :|| | ||  || :
Db        538 WAAAAAWATTWTTTTTTWWTMATAAAAAATTTAWTAWCATATTTATWTTATATAATATAW 479

Qy       1769 GAAC 1772
              | |
Db        478 AATC 475


RESULT 44
CNS01038/c
LOCUS       CNS01038                1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03O17 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098462
VERSION     AL098462.1  GI:5610073
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03O17"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             2.3%;  Score 46;  DB 10;  Length 1201;
  Best Local Similarity   32.6%;  Pred. No. 2.8;
  Matches   71;  Conservative   53;  Mismatches   94;  Indels    0;  Gaps    0;

Qy       1535 ATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCAC 1594
              | |: ::|: :|:| :|:: :    | ::|||     :| : |  ::|| | |  |
Db       1193 AAAWAWWAWTWAWTWWTWWWTWATTWTTHWATAWNTAWWTTWTTTWWWTTWTTTTTTTTT 1134

Qy       1595 TTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAAC 1654
              ||   || |  ||    | |   :|:  |: :|||| :|  || ||: :|:||: |   :
Db       1133 TTTATATTAAAAAAAAAATTWAWAWNAAWAWAAAAAWANTATTTAWAWAWTTMWATATWT 1074

Qy       1655 AAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTT 1714
                :|   |::  | : |  |:||: || ||||| |||  :: :     | : :|| |||
Db       1073 TTTHTATATWWWATWWTTTAAWAAWAAATAAAAAAAAAATWHYHATTTATAWTWATTTTT 1014

Qy       1715 GTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTT 1752
                |  | :  :|:||   ::| :::: ::::  :::|
Db       1013 TCTNTTTWWTTWCMTTWYWWMTTHMYWTMMMMHMHHWT 976


RESULT 45
CNS0161D/c
LOCUS       CNS0161D               1225 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106171
VERSION     AL106171.1  GI:5620504
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1225)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1225
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15C18"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             2.3%;  Score 46;  DB 10;  Length 1225;
  Best Local Similarity   33.0%;  Pred. No. 2.8;
  Matches  109;  Conservative   58;  Mismatches  163;  Indels    0;  Gaps    0;

Qy       1477 TTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCAT 1536
              |:| || |  : :   | |: |  :|: : |:| | || : |:||    | |      |:
Db       1218 TWATATTTTTWTKATTTTTWWAWAWAWWWWAWAWAAAAWWTTWTTTATAAAAATATATAW 1159

Qy       1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
              || ||    ::: | :|  ||   |  :||    :      :  | | |   : : :
Db       1158 ATWTATTATWWWAAAWAWTTTTATATWATTWTAWATAAAAWAAAAATATTAWWWAWWWA 1099

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
              |: |:|:  || :  ::   : |  |  |||:  :| |    |    : :| :: :| ||
Db       1098 AWAAWAWAWAAWWATWWTWTWKAATATAAAARTAWATATATAAAAATWAWAWWWTWAWAA 1039

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
                |  |||  |   : |  |:| || |||:| |:|  |  |  :: | ||:::
Db       1038 AATATATTTTTTTTTWTTTTAWAAAAAAAAWAAAWATAAATTATWWATAAAWWWTAAAAA 979

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATC 1776
               ||| |  |  || | |:| | ||:|||: |:|||   || ::| | |: |  : |
Db        978 AAAATAAWTTATTTWAWAWTTTTTTRAAAWWTWTTTTTTTTTWWTTTAAWAAWTWWTTAT 919

Qy       1777 TTGACTCAAGTGGACAGTGGTGAGCTTTTT 1806
              :|:  |  :  ::| | |  :    |||||
Db        918 WTRTTTWTRTRRRARAATWWWTTTTTTTTT 889


Search completed: January 22, 2006, 21:08:58
Job time : 5429.54 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 01:00:32 ; Search time 6808.48 Seconds
(without alignments)
16706.187 Million cell updates/sec

Title: US-09-728-552A-3_COPY_78595_80595
Perfect score: 2001
Sequence: 1 atctttatttgtatgagaaa.....aagtcattctgccagaattc 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	39678	6	AX033938	AX033938 Sequence
	2	2001	100.0	80622	8	AF222855	AF222855 Homo sapi
	3	1984.8	99.2	18443	6	AX033914	AX033914 Sequence
	4	1983.6	99.1	196934	8	AL355530	AL355530 Human DNA
	5	1895.2	94.7	39238	6	AX033939	AX033939 Sequence
	6	1895.2	94.7	80155	8	AF042484	AF042484 Homo sapi
	7	1895.2	94.7	80202	8	AF222854	AF222854 Homo sapi
c	8	589	29.4	71032	14	AC084084	AC084084 Homo sapi
c	9	407	20.3	258662	14	AC152318	AC152318 Bos tauru
	10	296	14.8	222814	9	AC134591	AC134591 Mus muscu
	11	263.4	13.2	237778	14	AC128358	AC128358 Rattus no
	12	97.6	4.9	300	6	AX033930	AX033930 Sequence
	13	60.8	3.0	7218	6	I66494	I66494 Sequence 14
	14	58.8	2.9	2000	6	AX655393	AX655393 Sequence
c	15	58.4	2.9	7218	6	I66494	I66494 Sequence 14
c	16	56.4	2.8	143324	8	AL137881	AL137881 Human DNA
	17	56.4	2.8	163554	14	AC021786	AC021786 Homo sapi
	18	55	2.7	193035	8	AC008180	AC008180 Homo sapi
	19	53.8	2.7	135653	8	AC130404	AC130404 Homo sapi
c	20	53.8	2.7	151751	14	AC023243	AC023243 Homo sapi
c	21	53.8	2.7	158059	14	AC041019	AC041019 Homo sapi
	22	53.8	2.7	187341	14	AC024980	AC024980 Homo sapi
	23	53.8	2.7	190310	14	AC023225	AC023225 Homo sapi
	24	52.8	2.6	250663	2	AE014826	AE014826 Plasmodiu
	25	52.2	2.6	167004	5	AL954840	AL954840 Zebrafish
c	26	51.8	2.6	188520	14	BX927412	BX927412 Danio rer
c	27	51.2	2.6	98286	8	AC092392	AC092392 Homo sapi
	28	51.2	2.6	101840	8	AC099062	AC099062 Homo sapi
c	29	51.2	2.6	253151	2	AE014842	AE014842 Plasmodiu
	30	50.6	2.5	123019	8	AC111006	AC111006 Homo sapi
c	31	49.8	2.5	191481	14	AC108123	AC108123 Homo sapi
	32	49.4	2.5	431	6	BD038954	BD038954 Sequence
	33	49.4	2.5	431	6	AX903421	AX903421 Sequence
c	34	49.4	2.5	1428	6	CS083840	CS083840 Sequence
c	35	49.4	2.5	155106	8	AC104069	AC104069 Homo sapi
	36	49.4	2.5	157671	8	AL450263	AL450263 Human DNA
c	37	49.4	2.5	161027	14	AL354705	AL354705 Homo sapi
c	38	49.4	2.5	349751	2	PFMAL4P3	AL035476 Plasmodiu
c	39	49.2	2.5	155357	14	CR405684	CR405684 Danio rer
c	40	49	2.4	214395	5	CR354428	CR354428 Zebrafish
c	41	48.4	2.4	113695	8	AC087447	AC087447 Homo sapi
c	42	48.4	2.4	176697	14	AC021170	AC021170 Homo sapi
c	43	48.4	2.4	190737	14	AC019143	AC019143 Homo sapi
c	44	48.4	2.4	194642	8	AC090163	AC090163 Homo sapi
	45	48.2	2.4	1818	1	AF267226	AF267226 Candidatu
c	46	48.2	2.4	157957	8	AL445590	AL445590 Human DNA
c	47	48.2	2.4	167254	8	CNS05TDS	AL357093 Human chr
c	48	48.2	2.4	200542	8	CNS01RG3	AL157736 Human chr
	49	48	2.4	146570	2	AC117076	AC117076 Dictyoste
c	50	47.8	2.4	118533	14	AC150245	AC150245 Medicago
	51	47.8	2.4	122151	5	AL935051	AL935051 Zebrafish
	52	47.8	2.4	126053	8	AL359916	AL359916 Human DNA
	53	47.4	2.4	110000	14	PFMAL8P1_04	Continuation (5 of
c	54	47.4	2.4	250743	2	AE014836	AE014836 Plasmodiu
	55	47.2	2.4	3706	2	PFAATUBII	M34390 P.falciparu
c	56	47	2.3	1141	6	AR579680	AR579680 Sequence
c	57	47	2.3	1141	6	AX083744	AX083744 Sequence
c	58	47	2.3	7536	2	AY392440	AY392440 Dictyoste
	59	47	2.3	153232	5	AL928821	AL928821 Zebrafish
	60	47	2.3	211406	5	BX004890	BX004890 Zebrafish
	61	47	2.3	218895	14	CR555304	CR555304 Danio rer
c	62	46.8	2.3	3854	8	BC022510	BC022510 Homo sapi
c	63	46.8	2.3	83951	8	AL354824	AL354824 Human DNA
c	64	46.8	2.3	129641	8	AC008560	AC008560 Homo sapi
c	65	46.8	2.3	138045	14	AC152477	AC152477 Dasypus n
	66	46.8	2.3	141119	14	AC027324	AC027324 Homo sapi
c	67	46.8	2.3	142650	8	AC112656	AC112656 Homo sapi
c	68	46.8	2.3	189203	14	AC019125	AC019125 Homo sapi
c	69	46.8	2.3	327650	1	MPULM01	AL445563 Mycoplasm
c	70	46.8	2.3	333321	2	AC116986	AC116986 Dictyoste
c	71	46.6	2.3	1597	8	AK130927	AK130927 Homo sapi
	72	46.6	2.3	69967	14	AC126337	AC126337 Homo sapi
	73	46.6	2.3	104637	8	AF186191	AF186191 Homo sapi
	74	46.6	2.3	123028	8	AC125239	AC125239 Homo sapi
	75	46.6	2.3	152405	8	AF186190	AF186190 Homo sapi
c	76	46.6	2.3	173540	14	AC125240	AC125240 Homo sapi
	77	46.6	2.3	182600	14	AC023088	AC023088 Homo sapi
c	78	46.6	2.3	252394	2	AE014833	AE014833 Plasmodiu
c	79	46.4	2.3	56479	8	AC006034	AC006034 Homo sapi
c	80	46.4	2.3	159250	8	AC073628	AC073628 Homo sapi
	81	46.4	2.3	160463	14	AC011566	AC011566 Homo sapi
	82	46.4	2.3	220579	8	AC130456	AC130456 Homo sapi
	83	46.4	2.3	220633	8	HUU91321	U91321 Human Chrom
c	84	46.4	2.3	233388	14	AC161519	AC161519 Mus muscu
	85	46.4	2.3	243763	14	AC095657	AC095657 Rattus no
c	86	46.4	2.3	246366	14	AC094648	AC094648 Rattus no
c	87	46.2	2.3	142924	14	CR762428	CR762428 Danio rer
c	88	46.2	2.3	158038	5	AL928875	AL928875 Zebrafish
	89	46.2	2.3	250793	14	CR388030	CR388030 Danio rer
	90	46	2.3	1141	6	AR579680	AR579680 Sequence
	91	46	2.3	1141	6	AX083744	AX083744 Sequence

	92	46	2.3	110000	2	AC116957_1 Continuation (2 of
c	93	46	2.3	145257	14	AC108178 AC108178 Felis cat
c	94	46	2.3	153796	14	AC157265 AC157265 Rhinoloph
	95	46	2.3	155848	8	AC104996 AC104996 Homo sapi
	96	46	2.3	164400	14	CR936374 CR936374 Danio rer
c	97	46	2.3	212314	14	AC151026 AC151026 Callithri
c	98	46	2.3	219112	14	AC164974 AC164974 Pan trogl
c	99	46	2.3	281723	2	PFA929359 AL929359 Plasmodiu
c	100	45.8	2.3	599	10	BV418118 BV418118 S229P6471

ALIGNMENTS

RESULT 1
AX033938
LOCUS AX033938 39678 bp DNA linear PAT 21-SEP-2000
DEFINITION Sequence 30 from Patent WO9851790.
ACCESSION AX033938
VERSION AX033938.1 GI:10280507
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified
unclassified sequences.
REFERENCE 1
AUTHORS Cancilla,M.R., Choo,K.H. and Du,S.D.
TITLE A novel nucleic acid molecule
JOURNAL Patent: WO 9851790-A 30 19-NOV-1998;
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES Location/Qualifiers
source 1. .39678
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN

Query Match 100.0%; Score 2001; DB 6; Length 39678;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37678 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 37737

Qy         61 TCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37738 TCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAA 37797

Qy        121 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37798 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAAC 37857

Qy        181 AAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37858 AAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCT 37917

Qy        241 GTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37918 GTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGA 37977

Qy        301 TGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37978 TGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTTT 38037

Qy        361 TTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38038 TTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAAC 38097

Qy        421 ATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38098 ATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTA 38157

Qy        481 CTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38158 CTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAG 38217

Qy        541 CTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38218 CTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTT 38277

Qy        601 TTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38278 TTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTT 38337

Qy        661 TTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38338 TTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCA 38397

Qy        721 GGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38398 GGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAG 38457

Qy        781 CATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38458 CATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTG 38517

Qy        841 ACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38518 ACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACAC 38577

Qy        901 GAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38578 GAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGCT 38637

Qy        961 GTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38638 GTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCAC 38697

Qy       1021 TAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38698 TAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGAC 38757

Qy       1081 CTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38758 CTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTC 38817

Qy       1141 CATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38818 CATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACA 38877

Qy       1201 GTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38878 GTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAAT 38937

Qy       1261 GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38938 GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCA 38997

Qy       1321 TAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAAT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38998 TAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAAT 39057

Qy       1381 ACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39058 ACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCTT 39117

Qy       1441 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39118 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 39177

Qy       1501 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39178 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 39237

Qy       1561 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 1620
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39238 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 39297

Qy       1621 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39298 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 39357

Qy       1681 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39358 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 39417

Qy       1741 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39418 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 39477

Qy       1801 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTAC 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39478 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTAC 39537

Qy       1861 AGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACT 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39538 AGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACT 39597

Qy       1921 TTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGT 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39598 TTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGT 39657

Qy       1981 GAAGTCATTCTGCCAGAATTC 2001
                |||||||||||||||||||||
Db      39658 GAAGTCATTCTGCCAGAATTC 39678
```

RESULT 2
AF222855
LOCUS AF222855 80622 bp DNA linear PRI 17-JUL-2000
DEFINITION Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION AF222855
VERSION AF222855.1 GI:9246845
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 80622)
AUTHORS Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M., Irvine,D.V., Saffery,R. and Choo,K.H.
TITLE The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification
JOURNAL Genome Res. 10 (6), 832-838 (2000)
PUBMED 10854414
REFERENCE 2 (bases 1 to 80622)
AUTHORS Barry,A.E.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd., Parkville,, Melbourne, Victoria 3052, Australia
REMARK Genomic sequence from human 10q25.2, clonelib=HC
FEATURES Location/Qualifiers
source 1. .80622
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q25.2"
/clone="HC"
repeat_region complement(3. .176)
/rpt_family="Alu"
repeat_region complement(979. .3564)
/rpt_family="Tigger1"
repeat_region 2967. .3094
/rpt_family="Alu"
repeat_region 3664. .3934
/rpt_family="Alu"
repeat_region complement(4920. .5181)
/rpt_family="Alu"
repeat_region complement(6287. .6579)
/rpt_family="Alu"
repeat_region 6625. .6966
/rpt_family="MLT2B2"
repeat_region complement(9215. .9490)
/rpt_family="Alu"
repeat_region complement(11280. .11565)
/rpt_family="Alu"
repeat_region complement(12428. .12675)
/rpt_family="Alu"
repeat_region complement(14304. .14583)
/rpt_family="Alu"
repeat_region 17503. .17794
/rpt_family="Alu"
repeat_region complement(20250. .20533)
/rpt_family="Alu"
repeat_region complement(23031. .23323)
/rpt_family="Alu"
repeat_region complement(24301. .24479)
/rpt_family="MIR"
repeat_region 24951. .25250
/rpt_family="Alu"
repeat_region complement(25987. .26342)
/rpt_family="THE1"
repeat_region 27097. .27335
/rpt_family="MER2"
repeat_region 27538. .27814
/rpt_family="Alu"
repeat_region 27891. .28161
/rpt_family="Alu"
repeat_region 29063. .29334
/rpt_family="Alu"
repeat_region complement(31025. .31093)
/rpt_family="MLT1"
repeat_region complement(31817. .32352)
/rpt_family="MER1"
repeat_region complement(32489. .32607)
/rpt_family="MLT1"
repeat_region 33039. .33316
/rpt_family="Alu"
repeat_region 33469. .33760
/rpt_family="Alu"
repeat_region complement(34527. .34590)
/rpt_family="MIR"
repeat_region 34689. .34995
/rpt_family="Alu"
repeat_region complement(36349. .36437)
/rpt_family="MIR"
repeat_region complement(38796. .39079)
/rpt_family="Alu"
repeat_region 39441. .39693
/rpt_family="Alu"
repeat_region complement(43176. .43237)
/rpt_family="MIR"
repeat_region complement(45203. .45478)
/rpt_family="Alu"
repeat_region 45892. .46025
/rpt_family="MIR"
repeat_region complement(46618. .46851)
/rpt_family="Alu"
repeat_region 48623. .48939
/rpt_family="Alu"
repeat_region complement(49048. .50698)
/rpt_family="L1"
repeat_region complement(50710. .50972)
/rpt_family="Alu"
repeat_region complement(51031. .51192)
/rpt_family="L1"
repeat_region complement(51450. .51739)

```
                     /rpt_family="Alu"
     repeat_region   complement(53270. .53562)
                     /rpt_family="Alu"
     repeat_region   54261. .54808
                     /rpt_family="LTR12"
     repeat_region   59460. .60054
                     /rpt_family="LTR12"
     repeat_region   60349. .60633
                     /rpt_family="Alu"
     repeat_region   complement(60669. .60801)
                     /rpt_family="MIR"
     repeat_region   complement(61684. .61852)
                     /rpt_family="MER5"
     repeat_region   63466. .63753
                     /rpt_family="Alu"
     repeat_region   64063. .64228
                     /rpt_family="MER5"
     repeat_region   64446. .64521
                     /rpt_family="MIR"
     repeat_region   65794. .65883
                     /rpt_family="MER5"
     repeat_region   66772. .66951
                     /rpt_family="MER20"
     repeat_region   complement(67895. .68416)
                     /rpt_family="MER41"
     repeat_region   complement(73192. .73470)
                     /rpt_family="Alu"
     repeat_region   74086. .74367
                     /rpt_family="Alu"
     repeat_region   75916. .76183
                     /rpt_family="Alu"
     repeat_region   complement(77771. .78059)
                     /rpt_family="Alu"
ORIGIN

  Query Match             100.0%;  Score 2001;  DB 8;  Length 80622;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78622 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 78681

Qy         61 TCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78682 TCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAA 78741

Qy        121 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78742 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAAC 78801

Qy        181 AAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78802 AAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCT 78861

Qy        241 GTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78862 GTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGA 78921

Qy        301 TGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78922 TGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTTT 78981

Qy        361 TTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78982 TTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAAC 79041

Qy        421 ATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79042 ATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTA 79101

Qy        481 CTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79102 CTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAG 79161

Qy        541 CTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79162 CTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTT 79221

Qy        601 TTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79222 TTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTT 79281

Qy        661 TTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79282 TTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCA 79341

Qy        721 GGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79342 GGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAG 79401

Qy        781 CATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79402 CATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTG 79461

Qy        841 ACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79462 ACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACAC 79521

Qy        901 GAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79522 GAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGCT 79581

Qy        961 GTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79582 GTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCAC 79641

Qy       1021 TAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79642 TAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGAC 79701

Qy       1081 CTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79702 CTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTC 79761

Qy       1141 CATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79762 CATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACA 79821

Qy       1201 GTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79822 GTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAAT 79881

Qy       1261 GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79882 GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCA 79941

Qy       1321 TAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAAT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79942 TAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAAT 80001

Qy       1381 ACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80002 ACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCTT 80061

Qy       1441 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80062 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 80121

Qy       1501 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80122 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 80181

Qy       1561 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80182 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 80241
```

```
Qy       1621 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80242 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 80301

Qy       1681 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80302 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 80361

Qy       1741 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80362 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 80421

Qy       1801 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTAC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80422 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTAC 80481

Qy       1861 AGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80482 AGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACT 80541

Qy       1921 TTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80542 TTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGT 80601

Qy       1981 GAAGTCATTCTGCCAGAATTC 2001
             |||||||||||||||||||||
Db      80602 GAAGTCATTCTGCCAGAATTC 80622
```

RESULT 3
AX033914
LOCUS AX033914 18443 bp DNA linear PAT 21-SEP-2000
DEFINITION Sequence 6 from Patent WO9851790.
ACCESSION AX033914
VERSION AX033914.1 GI:10280482
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified
unclassified sequences.
REFERENCE 1
AUTHORS Cancilla,M.R., Choo,K.H. and Du,S.D.
TITLE A novel nucleic acid molecule
JOURNAL Patent: WO 9851790-A 6 19-NOV-1998;
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES Location/Qualifiers
source 1. .18443
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN

Query Match 99.2%; Score 1984.8; DB 6; Length 18443;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1997; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

```
Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db       4292 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATTGT 4351

Qy         61 TCAGGA-CCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 119
              |||||  | || |||||||||||||||||||||||||||||||||||||||||||||||||
Db       4352 TCAGAACCCTTGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 4411

Qy        120 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4412 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 4471

Qy        180 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4472 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 4531

Qy        240 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4532 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 4591

Qy        300 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4592 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 4651

Qy        360 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4652 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 4711

Qy        420 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4712 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 4771

Qy        480 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4772 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 4831

Qy        540 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4832 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 4891

Qy        600 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4892 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 4951

Qy        660 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4952 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 5011

Qy        720 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5012 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 5071

Qy        780 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5072 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 5131

Qy        840 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5132 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 5191

Qy        900 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5192 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 5251

Qy        960 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5252 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 5311

Qy       1020 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5312 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 5371

Qy       1080 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5372 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 5431

Qy       1140 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5432 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 5491

Qy       1200 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5492 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 5551

Qy       1260 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5552 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 5611

Qy       1320 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 1379
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5612 ATAATTATTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 5671

Qy       1380 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5672 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 5731

Qy       1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5732 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 5791

Qy       1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5792 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 5851

Qy       1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5852 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 5911

Qy       1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5912 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 5971

Qy       1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 1739
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5972 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 6031

Qy       1740 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 1799
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6032 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 6091

Qy       1800 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 1859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6092 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 6151

Qy       1860 CAGGTGCCAAGTTAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGAC 1919
              |||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db       6152 CAGGTGCCAAGTTAGTGCCCAAGCTTGMTCTTAAAAGCATACTGGATTTTGTTTTAGAC 6211

Qy       1920 TTTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGG 1979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6212 TTTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGG 6271

Qy       1980 TGAAGTCATTCTGCCAGAATTC 2001
              ||||||||||||||||||||||
Db       6272 TGAAGTCATTCTGCCAGAATTC 6293
```

```
RESULT 4
AL355530
LOCUS       AL355530               196934 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-338L11 on chromosome 10 Contains
            the 5' end of the gene for novel gene (KIAA0534) and a CpG island,
            complete sequence.
ACCESSION   AL355530
VERSION     AL355530.6  GI:8894326
KEYWORDS    HTG; KIAA0534.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 196934)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 1, 2000 this sequence version replaced gi:8670617.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            RP11-338L11 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .196934
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-338L11"
                     /clone_lib="RPCI-11.2"
     misc_feature    1
                     /note="Clone_left_end: RP11-338L11"
     gene            join(68829. .69507,95654. .95737,97168. .97281,
                     103062. .103190,104794. .105002,135506. .135680,
                     141023. .141110,146500. .>146837)
                     /gene="RP11-537G20.1"
                     /locus_tag="RP11-537G20.1-002"
     gene            join(68829. .69100,72107. .72301,95654. .95737,97168. .97281,
                     103062. .103190,103876. .103977,104794. .105002,
                     135506. .135680,141023. .141110,146500. .146837)
                     /gene="RP11-537G20.1"
                     /locus_tag="RP11-537G20.1-003"
     mRNA            join(68829. .69507,95654. .95737,97168. .97281,
                     103062. .103190,104794. .105002,135506. .135680,
                     141023. .141110,146500. .>146837)
                     /gene="RP11-537G20.1"
                     /locus_tag="RP11-537G20.1-002"
                     /product="novel protein"
                     /note="match: cDNAs: BC035157.1"
     mRNA            join(68829. .69100,72107. .72301,95654. .95737,97168. .97281,
                     103062. .103190,103876. .103977,104794. .105002,
                     135506. .135680,141023. .141110,146500. .146837)
                     /gene="RP11-537G20.1"
                     /locus_tag="RP11-537G20.1-002"
                     /product="novel protein"
                     /note="match: cDNAs: BC029592.1"
     CDS             join(69215. .69507,95654. .95737,97168. .97281,
                     103062. .103190,104794. .105002,135506. .135680,
                     141023. .141110,146500. .>146837)
                     /gene="RP11-537G20.1"
                     /locus_tag="RP11-537G20.1-002"
                     /standard_name="OTTHUMP00000020556"
                     /codon_start=1
                     /product="novel protein"
                     /protein_id="CAI13482.1"
                     /db_xref="GI:55958328"
                     /db_xref="InterPro:IPR000859"
                     /db_xref="InterPro:IPR006209"
                     /db_xref="InterPro:IPR006210"
                     /db_xref="InterPro:IPR006652"
                     /db_xref="InterPro:IPR011498"
                     /db_xref="UniProt/TrEMBL:Q5T5Y9"
                     /translation="METGGRARTGTPQPAAPGVWRARPAGGGGGGASSWLLDGNSWLL
```

```
                CYGFLYLALYAQVSQSKPCERTGSCFSGRCVNSTCLCDPGWVGDQCQHCQGRFKLTEP
                SGYLTDGPINYKYKTKCTWLIEGYPNAVLRLRFNHFATECSWDHMYVYDGDSIYAPLI
                AVLSGLIVPEIRGNETVPEVVTTSGYALLHFFSDAAYNLTGFNIFYSINSCPNNCSGH
                GKCTTSVSVPSQVYCECDKYWKGEACDIPYCKANCGSPDHGYCDLTGEKLCVCNDSWQ
                GPDCSLNVPSTESYWILPNVKPFSPSVGRASHKAVLHGKFMWVIGGYTFNYSSFQMVL
                NYNLESSIWNVGTPSRGPLQRYGHSLALYQENIFMYGGRIETNDGNVTDELWVFNIHS
                QSWSTKTPTVLGHGQQYAVEGHSAHIMELDSRDVVMIIIFGYSAIYGYTSSIQEYHIC
                ELLKNCNFFIDWECFSLIKSSYEFSQ"
gene            join(<141072. .141110,146500. .146755,191160. .191343,
                AL512304.10:22131. .22285,AL512304.10:45441. .45525,
                AL512304.10:47045. .47253,AL512304.10:48574. .48692,
                AL512304.10:61636. .61768,AL512304.10:66497. .66678,
                AL357059.13:11678. .11891,AL357059.13:13499. .13687,
                AL357059.13:27162. .27380,AL357059.13:45926. .46063,
                AL356100.8:34176. .34269,AL356100.8:65767. .65819,
                AL356100.8:101458. .101552,AL356100.8:106691. .106769,
                AL356100.8:108689. .108846,AL356100.8:158780. .158841,
                AL356100.8:188975. .189053,AL392087.7:178400. .178507,
                AC091667.9:114936. .115050,AC022542.5:69593. .73927)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-001"
mRNA            join(<141072. .141110,146500. .146755,191160. .191343,
                AL512304.10:22131. .22285,AL512304.10:45441. .45525,
                AL512304.10:47045. .47253,AL512304.10:48574. .48692,
                AL512304.10:61636. .61768,AL512304.10:66497. .66678,
                AL357059.13:11678. .11891,AL357059.13:13499. .13687,
                AL357059.13:27162. .27380,AL357059.13:45926. .46063,
                AL356100.8:34176. .34269,AL356100.8:65767. .65819,
                AL356100.8:101458. .101552,AL356100.8:106691. .106769,
                AL356100.8:108689. .108846,AL356100.8:158780. .158841,
                AL356100.8:188975. .189053,AL392087.7:178400. .178507,
                AC091667.9:114936. .115050,AC022542.5:69593. .73927)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-001"
                /product="novel protein"
                /note="match: ESTs: AI588945.1 AI817331.1 R37016.1
                match: cDNAs: AB011106.2 BC027764.1"
CDS             join(<141072. .141110,146500. .146755,191160. .191343,
                AL512304.10:22131. .22285,AL512304.10:45441. .45525,
                AL512304.10:47045. .47253,AL512304.10:48574. .48692,
                AL512304.10:61636. .61768,AL512304.10:66497. .66678,
                AL357059.13:11678. .11891,AL357059.13:13499. .13687,
                AL357059.13:27162. .27380,AL357059.13:45926. .46063,
                AL356100.8:34176. .34269,AL356100.8:65767. .65819,
                AL356100.8:101458. .101552,AL356100.8:106691. .106769,
                AL356100.8:108689. .108846,AL356100.8:158780. .158841,
                AL356100.8:188975. .189053,AL392087.7:178400. .178507,
                AC091667.9:114936. .115050,AC022542.5:69593. .69714)
                /gene="RP11-537G20.1"
                /locus_tag="RP11-537G20.1-001"
                /standard_name="OTTHUMP00000020555"
                /codon_start=1
                /product="novel protein"
                /protein_id="CAI13483.1"
                /db_xref="GI:55958329"
                /db_xref="GOA:Q5VV63"
                /db_xref="InterPro:IPR001304"
                /db_xref="InterPro:IPR002049"
                /db_xref="InterPro:IPR002165"
                /db_xref="InterPro:IPR003659"
                /db_xref="InterPro:IPR006652"
                /db_xref="InterPro:IPR011498"
                /db_xref="UniProt/TrEMBL:Q5VV63"
                /translation="PLQRYGHSLALYQENIFMYGGRIETNDGNVTDELWVFNIHSQSW
                STKTPTVLGHGQQYAVEGHSAHIMELDSRDVVMIIIFGYSAIYGYTSSIQEYHISSNT
                WLVPETKGAIVQGGGYGHTSVYDEITKSIYVHGGYKALPGNKYGLVDDLYKYEVNTKTW
                TILKESGFARYLHSAVLINGAMLIFGGNTHNDTSLSNGAKCFSADFLAYDIACDEWKI
                LPKPNLHRDVNRFGHSAVVINGSMYIFGGFSSVLLNDILVYKPPNCKAFRDEELCKNA
                GPGIKCVWNKNHCESWESGNTNNILRAKCPPKTAASDDRCYRYADCASCTANTNGCQW
                CDDKKCISANSNCSMSVKNYTKCHVRNEQICNKLTSCKSCSLNLNCQWDQRQQECQAL
                PAHLCGEGWSHIGDACLRVNSSRENYDNAKLYCYNLSGNLASLTTSKEVEFVLDEIQK
                YTQQKVSPWVGLRKINISYWGWEDMSPFTNTTLQWLPGEPNDSGFCAYLERAAVAGLK
                ANPCTSMANGLVCEKPVVSPNQNARPCKKPCSLRTSCSNCTSNGMECMWCSSTKRCVD
                SNAYIISFPYGQCLEWQTATCSPQNCSGLRTCGQCLEQPGCGWCNDPSNTGRGHCIEG
                SSRGPMKLIGMHHSEMVLDTNLCPKEKNYEWSFIQCPACQCNGHSTCINNNVCEQCKN
                LTTGKQCQDCMPGYYGDPTNGGQCTACTCSGHANICHLHTGKCFCTTKGIKGDQCQLC
                DSENRYVGNPLRGTCYYSLLIDYQFTFSLLQEDDRHHTAINFIANPEQSNKNLDISIN
                ASNNFNLNITWSVGSTAGTISGEETSIVSKNNIKEYRDSFSYEKFNFRSNPNITFYVY
                VSNFSWPIKIQIAFSQHNTIMDLVQFFVTFFSCFLSLLLVAAVVWKIKQTCWASRRRE
                QLLRERQQMASRPFASVDVALEVGAEQTEFLRGPLEGAPKPIAIEPCAGNRAAVLTVF
                LCLPRGSSGAPPPGQSGLAIASALIDISQQKASDSKDKTSGVRNRKHLSTRQGTCV"
     misc_feature    196934
                /note="Clone_right_end: RP11-338L11"
ORIGIN

  Query Match             99.1%;  Score 1983.6;  DB 8;  Length 196934;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1997;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      36800 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATTGT 36859

Qy         61 TCAGGA-CCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 119
              |||| | || ||||||||||||||||||||||||||||||||||||||||||||||||
Db      36860 TCAGAACCCTTGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 36919

Qy        120 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36920 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 36979

Qy        180 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36980 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 37039

Qy        240 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37040 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 37099

Qy        300 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37100 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 37159

Qy        360 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37160 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 37219

Qy        420 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37220 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 37279

Qy        480 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37280 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 37339

Qy        540 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37340 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 37399

Qy        600 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37400 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 37459

Qy        660 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37460 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 37519

Qy        720 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37520 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 37579

Qy        780 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      37580 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 37639

Qy        840 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 899
```

```
Db      37640 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 37699

Qy        900 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 959

Db      37700 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAAAGC 37759

Qy        960 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 1019

Db      37760 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 37819

Qy       1020 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 1079

Db      37820 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 37879

Qy       1080 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 1139

Db      37880 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 37939

Qy       1140 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 1199

Db      37940 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 37999

Qy       1200 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 1259

Db      38000 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 38059

Qy       1260 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 1319

Db      38060 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 38119

Qy       1320 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 1379

Db      38120 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 38179

Qy       1380 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 1439

Db      38180 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 38239

Qy       1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499

Db      38240 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 38299

Qy       1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559

Db      38300 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 38359

Qy       1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619

Db      38360 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 38419

Qy       1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679

Db      38420 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 38479

Qy       1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 1739

Db      38480 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 38539

Qy       1740 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 1799

Db      38540 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 38599

Qy       1800 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 1859

Db      38600 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 38659

Qy       1860 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGAC 1919

Db      38660 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGAC 38719

Qy       1920 TTTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGG 1979

Db      38720 TTTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGG 38779

Qy       1980 TGAAGTCATTCTGCCAGAATTC 2001

Db      38780 TGAAGTCATTCTGCCAGAATTC 38801
```

RESULT 5
AX033939
LOCUS AX033939 39238 bp DNA linear PAT 21-SEP-2000
DEFINITION Sequence 31 from Patent WO9851790.
ACCESSION AX033939
VERSION AX033939.1 GI:10280508
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified
unclassified sequences.
REFERENCE 1
AUTHORS Cancilla,M.R., Choo,K.H. and Du,S.D.
TITLE A novel nucleic acid molecule
JOURNAL Patent: WO 9851790-A 31 19-NOV-1998;
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES Location/Qualifiers
source 1. .39238
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN

Query Match 94.7%; Score 1895.2; DB 6; Length 39238;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1908; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60

Db      37327 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATTGT 37386

Qy         61 TCAGGA-CCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 119

Db      37387 TCAGAACCCTTGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 37446

Qy        120 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 179

Db      37447 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 37506

Qy        180 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 239

Db      37507 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 37566

Qy        240 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 299

Db      37567 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 37626

Qy        300 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 359

Db      37627 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 37686

Qy        360 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 419

Db      37687 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 37746

Qy        420 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 479

Db      37747 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 37806

Qy        480 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 539

Db      37807 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 37866

Qy        540 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 599

Db      37867 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 37926
```

```
Qy       600 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 659
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     37927 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 37986

Qy       660 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 719
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     37987 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 38046

Qy       720 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 779
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38047 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 38106

Qy       780 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 839
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38107 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 38166

Qy       840 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 899
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38167 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 38226

Qy       900 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 959
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38227 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 38286

Qy       960 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 1019
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38287 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 38346

Qy      1020 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 1079
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38347 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 38406

Qy      1080 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 1139
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38407 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 38466

Qy      1140 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 1199
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38467 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 38526

Qy      1200 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 1259
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38527 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 38586

Qy      1260 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 1319
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38587 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 38646

Qy      1320 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 1379
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38647 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 38706

Qy      1380 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 1439
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38707 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 38766

Qy      1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38767 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 38826

Qy      1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38827 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 38886

Qy      1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38887 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 38946

Qy      1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     38947 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 39006

Qy      1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 1739
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     39007 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 39066

Qy      1740 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 1799
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     39067 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 39126

Qy      1800 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 1859
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     39127 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 39186

Qy      1860 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 1911
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     39187 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 39238
```

```
RESULT 6
AF042484
LOCUS       AF042484               80155 bp    DNA     linear   PRI 22-JUN-2000
DEFINITION  Homo sapiens sequence of the 80 kb core centromere protein binding
            region of the mar del(10) neocentromere corresponding to human
            chromosome 10q25.2, complete sequence.
ACCESSION   AF042484
VERSION     AF042484.1  GI:4205782
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,K.H.
  TITLE     Sequence analysis of an 80 kb human neocentromere
  JOURNAL   Hum. Mol. Genet. 8 (2), 217-227 (1999)
   PUBMED   9931329
REFERENCE   2  (bases 1 to 80155)
  AUTHORS   Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville, Melbourne, Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .80155
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /note="Sequence information obtained from BAC clones
                     prepared from cell line BE2C1-18-5f"
     repeat_region   complement(join(49. .525,3628. .4545,4849. .5053))
                     /rpt_family="L1MEc"
                     /rpt_type=dispersed
     repeat_region   join(560. .597,3184. .3281)
                     /rpt_family="MER58"
                     /rpt_type=dispersed
     repeat_region   complement(join(598. .2553,2716. .3183))
                     /rpt_family="TIGGER1"
                     /rpt_type=dispersed
     repeat_region   2558. .2581
                     /standard_name="T1"
                     /note="100% homology"
                     /rpt_type=direct
                     /rpt_unit="(caggcacagtgg)2"
     repeat_region   2570. .2715
                     /rpt_family="AluSg/x"
                     /rpt_type=dispersed
     repeat_region   3021. .3050
                     /standard_name="T2"
                     /note="83.3% homology"
                     /rpt_type=direct
                     /rpt_unit="(taacaaagtg)3"
```

```
repeat_region   3283. .3558
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   4173. .4211
                /standard_name="T3"
                /note="82.1% homology"
                /rpt_type=direct
                /rpt_unit="(aaaaaaataattt)3"
repeat_region   complement(4546. .4848)
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   complement(5909. .6208)
                /rpt_family="AluSg"
                /rpt_type=dispersed
repeat_region   join(6232. .6573,6638. .6688)
                /rpt_family="MLT2CB"
                /rpt_type=dispersed
repeat_region   6598. .6645
                /standard_name="AC24"
                /note="100% homology"
                /rpt_type=tandem
                /rpt_unit="ac"
repeat_region   6598. .6637
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="ca"
repeat_region   7091. .7134
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   7808. .7911
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   8349. .8524
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   complement(8839. .9124)
                /rpt_family="AluJb"
                /rpt_type=dispersed
repeat_region   9180. .9280
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   10070. .10170
                /note="low complexity"
                /rpt_family="purine-rich"
repeat_region   10097. .10136
                /standard_name="GA1"
                /note="82.5% homology"
                /rpt_type=direct
                /rpt_unit="(gaaagaaagg)4"
repeat_region   10175. .10339
                /rpt_family="FRAM"
                /rpt_type=dispersed
repeat_region   complement(10899. .11197)
                /rpt_family="AluSx"
                /rpt_type=dispersed
misc_feature    join(11627. .11757,12434. .12512,13443. .13562)
                /note=">98% identity"
                /db_xref="dbEST:AA774571"
repeat_region   11996. .12028
                /standard_name="AT4"
                /note="100% homology"
                /rpt_type=tandem
                /rpt_unit="(attt)8"
repeat_region   complement(12010. .12312)
                /rpt_family="AluJb"
                /rpt_type=dispersed
repeat_region   12327. .12433
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(13946. .14235)
                /rpt_family="AluSq"
                /rpt_type=dispersed
repeat_region   15130. .15159
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   15164. .15753
                /standard_name="AT28"
                /note="82.1% homology"
                /rpt_type=direct
                /rpt_unit="(atgtatatatgtgtatatagacataaat)21"
repeat_region   join(15170. .15289,15410. .15612)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="ta"
repeat_region   15617. .15731
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="cata"
repeat_region   16004. .16100
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16128. .16364)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16407. .16592)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(16657. .16889)
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   16914. .16936
                /note="low complexity"
                /rpt_family="AT-rich"
repeat_region   17102. .17386
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   17406. .17564
                /standard_name="GA2"
                /note="81.8% homology"
                /rpt_type=direct
repeat_region   17416. .17543
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   17647. .17702
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   complement(19863. .20161)
                /rpt_family="AluSx"
                /rpt_type=dispersed
repeat_region   complement(20842. .20913)
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   complement(21366. .21414)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="gaaa"
repeat_region   21375. .21639
                /standard_name="CT1"
                /note="75% homology"
                /rpt_type=dispersed
repeat_region   complement(21415. .21536)
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="tggaa"
repeat_region   complement(21524. .21646)
                /note="low complexity"
                /rpt_family="purine-rich"
repeat_region   22409. .22646
                /rpt_family="L1M4"
                /rpt_type=dispersed
repeat_region   22560. .22604
                /standard_name="T4"
                /note="80% homology"
                /rpt_type=direct
                /rpt_unit="(aatattacaataatt)3"
```

```
Query Match             94.7%;  Score 1895.2;  DB 8;  Length 80155;
Best Local Similarity   99.8%;  Pred. No. 0;
Matches 1908;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      78244 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATTGT 78303

Qy         61 TCAGGA-CCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 119
              |||| | || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78304 TCAGAACCCTTGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 78363

Qy        120 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78364 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 78423

Qy        180 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78424 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 78483

Qy        240 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78484 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 78543

Qy        300 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78544 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 78603

Qy        360 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78604 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 78663

Qy        420 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78664 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 78723

Qy        480 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78724 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 78783

Qy        540 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78784 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 78843

Qy        600 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78844 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 78903

Qy        660 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78904 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 78963

Qy        720 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78964 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 79023

Qy        780 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79024 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 79083

Qy        840 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79084 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 79143

Qy        900 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79144 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 79203

Qy        960 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79204 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 79263

Qy       1020 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79264 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 79323

Qy       1080 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79324 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 79383

Qy       1140 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79384 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 79443

Qy       1200 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79444 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 79503

Qy       1260 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79504 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 79563

Qy       1320 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79564 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 79623

Qy       1380 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79624 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 79683

Qy       1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79684 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 79743

Qy       1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79744 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 79803

Qy       1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79804 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 79863

Qy       1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79864 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 79923

Qy       1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 1739
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79924 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 79983

Qy       1740 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 1799
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79984 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 80043

Qy       1800 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 1859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80044 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 80103

Qy       1860 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 1911
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80104 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 80155
```

```
RESULT 7
AF222854
LOCUS       AF222854               80202 bp    DNA     linear   PRI 17-JUL-2000
DEFINITION  Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION   AF222854
VERSION     AF222854.1  GI:9246844
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 80202)
  AUTHORS   Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
            Irvine,D.V., Saffery,R. and Choo,K.H.
  TITLE     The 10q25 neocentromere and its inactive progenitor have identical
            primary nucleotide sequence: further evidence for epigenetic
            modification
  JOURNAL   Genome Res. 10 (6), 832-838 (2000)
   PUBMED   10854414
REFERENCE   2  (bases 1 to 80202)
  AUTHORS   Barry,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
            Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
            Parkville,, Melbourne, Victoria 3052, Australia
  REMARK    Human genomic sequence from 10q25.2, clonelib=NC, second release
FEATURES             Location/Qualifiers
     source          1. .80202
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25.2"
                     /clone="NC"
     repeat_region   complement(598. .3183)
                     /rpt_family="Tigger1"
     repeat_region   2587. .2715
                     /rpt_family="Alu"
     repeat_region   3283. .3553
                     /rpt_family="Alu"
     repeat_region   complement(4548. .4809)
                     /rpt_family="Alu"
     repeat_region   complement(5919. .6211)
                     /rpt_family="Alu"
     repeat_region   6257. .6599
                     /rpt_family="MLT2B2"
     repeat_region   complement(8848. .9123)
                     /rpt_family="Alu"
     repeat_region   complement(10913. .11198)
                     /rpt_family="Alu"
     repeat_region   complement(12062. .12313)
                     /rpt_family="Alu"
     repeat_region   complement(13957. .14236)
                     /rpt_family="Alu"
     repeat_region   17133. .17424
                     /rpt_family="Alu"
     repeat_region   complement(19882. .20197)
                     /rpt_family="Alu"
     repeat_region   complement(22664. .22956)
                     /rpt_family="Alu"
     repeat_region   complement(23935. .24113)
                     /rpt_family="MIR"
     repeat_region   24585. .24884
                     /rpt_family="Alu"
     repeat_region   complement(25618. .25973)
                     /rpt_family="THE1"
     repeat_region   26732. .26970
                     /rpt_family="MER2"
     repeat_region   27173. .27449
                     /rpt_family="Alu"
     repeat_region   27526. .27796
                     /rpt_family="Alu"
     repeat_region   28702. .28978
                     /rpt_family="Alu"
     repeat_region   complement(30668. .30736)
                     /rpt_family="MLT1"
     repeat_region   complement(31460. .31993)
                     /rpt_family="MER1"
     repeat_region   complement(32130. .32248)
                     /rpt_family="MLT1"
     repeat_region   32685. .32961
                     /rpt_family="Alu"
     repeat_region   33113. .33403
                     /rpt_family="Alu"
     repeat_region   complement(34169. .34233)
                     /rpt_family="MIR"
     repeat_region   34332. .34639
                     /rpt_family="Alu"
     repeat_region   complement(35993. .36081)
                     /rpt_family="MIR"
     repeat_region   complement(38438. .38721)
                     /rpt_family="Alu"
     repeat_region   39081. .39333
                     /rpt_family="Alu"
     repeat_region   complement(42817. .42877)
                     /rpt_family="MIR"
     repeat_region   complement(44845. .45119)
                     /rpt_family="Alu"
     repeat_region   45532. .45665
                     /rpt_family="MIR"
     repeat_region   complement(46257. .46490)
                     /rpt_family="Alu"
     repeat_region   48262. .48578
                     /rpt_family="Alu"
     repeat_region   complement(48687. .50337)
                     /rpt_family="L1"
     repeat_region   complement(50349. .50611)
                     /rpt_family="Alu"
     repeat_region   complement(50670. .50831)
                     /rpt_family="L1"
     repeat_region   complement(51085. .51374)
                     /rpt_family="Alu"
     repeat_region   complement(52905. .53197)
                     /rpt_family="Alu"
     repeat_region   53896. .54467
                     /rpt_family="LTR12"
     repeat_region   59129. .59719
                     /rpt_family="LTR12"
     repeat_region   60018. .60302
                     /rpt_family="Alu"
     repeat_region   complement(60338. .60470)
                     /rpt_family="MIR"
     repeat_region   complement(61353. .61521)
                     /rpt_family="MER5"
     repeat_region   63135. .63416
                     /rpt_family="Alu"
     repeat_region   63724. .63889
                     /rpt_family="MER5"
     repeat_region   64107. .64182
                     /rpt_family="MIR"
     repeat_region   65459. .65548
                     /rpt_family="MER5"
     repeat_region   66439. .66618
                     /rpt_family="MER20"
     repeat_region   complement(67562. .68084)
                     /rpt_family="MER41"
     repeat_region   complement(72861. .73139)
                     /rpt_family="Alu"
     repeat_region   73756. .74037
                     /rpt_family="Alu"
     repeat_region   75588. .75853
                     /rpt_family="Alu"
     repeat_region   complement(77441. .77729)
                     /rpt_family="Alu"
ORIGIN

  Query Match             94.7%;  Score 1895.2;  DB 8;  Length 80202;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1908;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      78291 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATTGT 78350

Qy         61 TCAGGA-CCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 119
              |||| | || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78351 TCAGAACCCTTGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 78410
```

```
Qy       120 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 179
Db     78411 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 78470

Qy       180 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 239
Db     78471 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 78530

Qy       240 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 299
Db     78531 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 78590

Qy       300 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 359
Db     78591 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 78650

Qy       360 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 419
Db     78651 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 78710

Qy       420 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 479
Db     78711 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 78770

Qy       480 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 539
Db     78771 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 78830

Qy       540 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 599
Db     78831 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 78890

Qy       600 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 659
Db     78891 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 78950

Qy       660 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 719
Db     78951 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 79010

Qy       720 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 779
Db     79011 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 79070

Qy       780 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 839
Db     79071 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 79130

Qy       840 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 899
Db     79131 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 79190

Qy       900 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 959
Db     79191 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 79250

Qy       960 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 1019
Db     79251 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 79310

Qy      1020 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 1079
Db     79311 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 79370

Qy      1080 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 1139
Db     79371 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 79430

Qy      1140 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 1199
Db     79431 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 79490

Qy      1200 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 1259
Db     79491 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 79550

Qy      1260 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 1319
Db     79551 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 79610

Qy      1320 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 1379
Db     79611 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 79670

Qy      1380 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 1439
Db     79671 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 79730

Qy      1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499
Db     79731 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 79790

Qy      1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559
Db     79791 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 79850

Qy      1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619
Db     79851 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 79910

Qy      1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679
Db     79911 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 79970

Qy      1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 1739
Db     79971 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 80030

Qy      1740 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 1799
Db     80031 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 80090

Qy      1800 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 1859
Db     80091 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 80150

Qy      1860 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 1911
Db     80151 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 80202
```

```
RESULT 8
AC084084/c
LOCUS       AC084084               71032 bp    DNA     linear   HTG 25-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-314C19 map 8, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC084084
VERSION     AC084084.2  GI:13446278
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 71032)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-314C19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71032)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
```

Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (12-OCT-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Mar 25, 2001 this sequence version replaced gi:10799449.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L11327
Center clone name: 314_C_19
----------.
* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 720: contig of 720 bp in length
* 721 820: gap of 100 bp
* 821 1557: contig of 737 bp in length
* 1558 1657: gap of 100 bp
* 1658 2404: contig of 747 bp in length
* 2405 2504: gap of 100 bp
* 2505 3299: contig of 795 bp in length
* 3300 3399: gap of 100 bp
* 3400 4153: contig of 754 bp in length
* 4154 4253: gap of 100 bp
* 4254 4978: contig of 725 bp in length
* 4979 5078: gap of 100 bp
* 5079 5815: contig of 737 bp in length
* 5816 5915: gap of 100 bp
* 5916 6648: contig of 733 bp in length
* 6649 6748: gap of 100 bp
* 6749 7520: contig of 772 bp in length
* 7521 7620: gap of 100 bp
* 7621 8347: contig of 727 bp in length
* 8348 8447: gap of 100 bp
* 8448 9227: contig of 780 bp in length
* 9228 9327: gap of 100 bp
* 9328 10052: contig of 725 bp in length
* 10053 10152: gap of 100 bp
* 10153 10869: contig of 717 bp in length
* 10870 10969: gap of 100 bp
* 10970 11718: contig of 749 bp in length
* 11719 11818: gap of 100 bp
* 11819 12557: contig of 739 bp in length
* 12558 12657: gap of 100 bp
* 12658 13412: contig of 755 bp in length
* 13413 13512: gap of 100 bp
* 13513 14255: contig of 743 bp in length
* 14256 14355: gap of 100 bp
* 14356 15086: contig of 731 bp in length
* 15087 15186: gap of 100 bp
* 15187 15901: contig of 715 bp in length
* 15902 16001: gap of 100 bp
* 16002 16711: contig of 710 bp in length
* 16712 16811: gap of 100 bp
* 16812 17539: contig of 728 bp in length
* 17540 17639: gap of 100 bp
* 17640 18351: contig of 712 bp in length
* 18352 18451: gap of 100 bp
* 18452 19162: contig of 711 bp in length
* 19163 19262: gap of 100 bp
* 19263 20054: contig of 792 bp in length
* 20055 20154: gap of 100 bp
* 20155 20891: contig of 737 bp in length
* 20892 20991: gap of 100 bp
* 20992 21720: contig of 729 bp in length
* 21721 21820: gap of 100 bp
* 21821 22553: contig of 733 bp in length
* 22554 22653: gap of 100 bp
* 22654 23420: contig of 767 bp in length
* 23421 23520: gap of 100 bp
* 23521 24234: contig of 714 bp in length
* 24235 24334: gap of 100 bp
* 24335 25069: contig of 735 bp in length
* 25070 25169: gap of 100 bp
* 25170 25915: contig of 746 bp in length
* 25916 26015: gap of 100 bp
* 26016 26751: contig of 736 bp in length
* 26752 26851: gap of 100 bp
* 26852 27574: contig of 723 bp in length
* 27575 27674: gap of 100 bp
* 27675 28437: contig of 763 bp in length
* 28438 28537: gap of 100 bp
* 28538 29258: contig of 721 bp in length
* 29259 29358: gap of 100 bp
* 29359 30068: contig of 710 bp in length
* 30069 30168: gap of 100 bp
* 30169 30918: contig of 750 bp in length
* 30919 31018: gap of 100 bp
* 31019 31770: contig of 752 bp in length
* 31771 31870: gap of 100 bp
* 31871 32658: contig of 788 bp in length
* 32659 32758: gap of 100 bp
* 32759 33501: contig of 743 bp in length
* 33502 33601: gap of 100 bp
* 33602 34347: contig of 746 bp in length
* 34348 34447: gap of 100 bp
* 34448 35165: contig of 718 bp in length
* 35166 35265: gap of 100 bp
* 35266 35998: contig of 733 bp in length
* 35999 36098: gap of 100 bp
* 36099 36802: contig of 704 bp in length
* 36803 36902: gap of 100 bp
* 36903 37632: contig of 730 bp in length
* 37633 37732: gap of 100 bp
* 37733 38454: contig of 722 bp in length
* 38455 38554: gap of 100 bp
* 38555 39275: contig of 721 bp in length
* 39276 39375: gap of 100 bp
* 39376 40113: contig of 738 bp in length
* 40114 40213: gap of 100 bp
* 40214 40948: contig of 735 bp in length
* 40949 41048: gap of 100 bp
* 41049 41792: contig of 744 bp in length
* 41793 41892: gap of 100 bp
* 41893 42630: contig of 738 bp in length
* 42631 42730: gap of 100 bp
* 42731 43459: contig of 729 bp in length
* 43460 43559: gap of 100 bp
* 43560 44277: contig of 718 bp in length

```
            *    44278    44377: gap of 100 bp
            *    44378    45120: contig of 743 bp in length
            *    45121    45220: gap of 100 bp
            *    45221    45942: contig of 722 bp in length
            *    45943    46042: gap of 100 bp
            *    46043    46775: contig of 733 bp in length
            *    46776    46875: gap of 100 bp
            *    46876    47609: contig of 734 bp in length
            *    47610    47709: gap of 100 bp
            *    47710    48446: contig of 737 bp in length
            *    48447    48546: gap of 100 bp
            *    48547    49285: contig of 739 bp in length
            *    49286    49385: gap of 100 bp
            *    49386    50178: contig of 793 bp in length
            *    50179    50278: gap of 100 bp
            *    50279    51008: contig of 730 bp in length
            *    51009    51108: gap of 100 bp
            *    51109    51846: contig of 738 bp in length
            *    51847    51946: gap of 100 bp
            *    51947    52672: contig of 726 bp in length
            *    52673    52772: gap of 100 bp
            *    52773    53505: contig of 733 bp in length
            *    53506    53605: gap of 100 bp
            *    53606    54330: contig of 725 bp in length
            *    54331    54430: gap of 100 bp
            *    54431    55166: contig of 736 bp in length
            *    55167    55266: gap of 100 bp
            *    55267    55979: contig of 713 bp in length
            *    55980    56079: gap of 100 bp
            *    56080    56823: contig of 744 bp in length
            *    56824    56923: gap of 100 bp
            *    56924    57626: contig of 703 bp in length

  Query Match             29.4%;  Score 589;  DB 14;  Length 71032;
  Best Local Similarity   84.1%;  Pred. No. 2.7e-119;
  Matches  650;  Conservative    0;  Mismatches  118;  Indels    5;  Gaps    5;

Qy       1048 CTCAGGCAGTGCAGCCTCAGGAGTGGAACTGACCTCTTATGTGTGTCCAGCCTTTCTTCC 1107
              |      | |  |  || |    |
Db      19287 CNNNNNCNGCACCCCCCCTNCCGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 19228

Qy       1108 TTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTCCATAGGAACATCAGTCCTGAATCCTCA 1167

Db      19227 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 19168

Qy       1168 GACCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGGATC 1227
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      19167 NNNNNCCATCTGGAGTAGTAAGTGCTCCTGACAGTCNTAGAAGTTGTCTACCGCTGGATC 19108

Qy       1228 T-CCAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAAT 1286
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19107 TCCCAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAAT 19048

Qy       1287 CTTGCTTTTTCACAAG-CCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAAT 1345
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      19047 CTTGCTTTTTCACAAGCCCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAAT 18988

Qy       1346 TTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTT 1405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18987 TTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTT 18928

Qy       1406 GCTAATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTA 1465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18927 GCTAATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTA 18868

Qy       1466 AGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCC 1525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18867 AGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCC 18808

Qy       1526 ACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTC 1585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18807 ACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTC 18748

Qy       1586 TCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAAT 1645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18747 TCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAAT 18688

Qy       1646 TAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTC 1705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18687 TAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTC 18628

Qy       1706 AAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGA 1765
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db      18627 AAAATATTTGTTTAATCAGGTCA-TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGA 18569

Qy       1766 TGAGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCC 1818
              |||||||||||||| |||||||||||||||||| ||||||||||||||||
Db      18568 TGAGAACTATCTT-ACTCAAGTGGACAGTGGTGAGC-TTTTGGCCTGTGGTCC 18518


RESULT 9
AC152318/c
LOCUS       AC152318              258662 bp    DNA     linear   HTG 01-JUL-2005
DEFINITION  Bos taurus clone CH240-2L12, *** SEQUENCING IN PROGRESS ***, 30
            unordered pieces.
ACCESSION   AC152318
VERSION     AC152318.3  GI:68227238
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 258662)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
```

```
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 258662)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2004) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 258662)
  AUTHORS   Cow Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 26, 2005 this sequence version replaced gi:58038134.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: FABV
                Center clone name: CH240-2L12
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 214934 bases at least Q40
                Consensus quality: 217760 bases at least Q30
                Consensus quality: 220382 bases at least Q20
                Estimated insert size: 219583; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3527: contig of 3527 bp in length
            *     3528     3577: gap of 50 bp
            *     3578     5414: contig of 1837 bp in length
            *     5415     5464: gap of 50 bp
            *     5465     6778: contig of 1314 bp in length
            *     6779     8809: gap of 2031 bp
            *     8810     9920: contig of 1111 bp in length
            *     9921    10020: gap of unknown length
            *    10021    12558: contig of 2538 bp in length
            *    12559    12608: gap of 50 bp
            *    12609    17883: contig of 5275 bp in length
            *    17884    17983: gap of unknown length
            *    17984    30723: contig of 12740 bp in length
            *    30724    30773: gap of 50 bp
            *    30774    46311: contig of 15538 bp in length
            *    46312    46361: gap of 50 bp
            *    46362    54121: contig of 7760 bp in length
            *    54122    54171: gap of 50 bp
            *    54172    69599: contig of 15428 bp in length
            *    69600    69649: gap of 50 bp
            *    69650    76462: contig of 6813 bp in length
            *    76463    76512: gap of 50 bp
            *    76513    78033: contig of 1521 bp in length
            *    78034    78133: gap of unknown length
            *    78134    81273: contig of 3140 bp in length
            *    81274    81323: gap of 50 bp
            *    81324    92153: contig of 10830 bp in length
            *    92154    92253: gap of unknown length
            *    92254    93265: contig of 1012 bp in length
            *    93266    94230: gap of 965 bp
            *    94231   115783: contig of 21553 bp in length
            *   115784   115833: gap of 50 bp
            *   115834   118884: contig of 3051 bp in length
            *   118885   118934: gap of 50 bp
            *   118935   151000: contig of 32066 bp in length
            *   151001   151050: gap of 50 bp
            *   151051   152191: contig of 1141 bp in length
            *   152192   182104: gap of 29913 bp
            *   182105   189192: contig of 7088 bp in length
            *   189193   189242: gap of 50 bp
            *   189243   192962: contig of 3720 bp in length
            *   192963   193012: gap of 50 bp
            *   193013   209262: contig of 16250 bp in length
            *   209263   209312: gap of 50 bp
            *   209313   231496: contig of 22184 bp in length
            *   231497   231767: gap of 271 bp
            *   231768   233967: contig of 2200 bp in length
            *   233968   234067: gap of unknown length
            *   234068   235110: contig of 1043 bp in length
            *   235111   235210: gap of unknown length
            *   235211   236716: contig of 1506 bp in length
            *   236717   236816: gap of unknown length
            *   236817   238176: contig of 1360 bp in length
            *   238177   238276: gap of unknown length
            *   238277   240200: contig of 1924 bp in length
            *   240201   240300: gap of unknown length
            *   240301   242112: contig of 1812 bp in length
            *   242113   242212: gap of unknown length
            *   242213   258662: contig of 16450 bp in length.
FEATURES             Location/Qualifiers
     source          1. .258662
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="CH240-2L12"
     gap             3528. .3577
                     /estimated_length=50
     gap             5415. .5464
                     /estimated_length=50
     gap             6779. .8809
                     /estimated_length=2031
     gap             9921. .10020
                     /estimated_length=unknown
     gap             12559. .12608
                     /estimated_length=50
     gap             17884. .17983
                     /estimated_length=unknown
     gap             30724. .30773
                     /estimated_length=50
     gap             46312. .46361
                     /estimated_length=50
     gap             54122. .54171
                     /estimated_length=50
     gap             69600. .69649
                     /estimated_length=50
     gap             76463. .76512
```

```
  Query Match             20.3%;  Score 407;  DB 14;  Length 258662;
  Best Local Similarity   66.9%;  Pred. No. 2.9e-79;
  Matches  814;  Conservative    0;  Mismatches  240;  Indels  163;  Gaps   10;

Qy        935 CTCAATACAGGATGTTCTAGAGAGCTGTATTTTGGGCACTTAACTATTCTCC-------A 987
              || |||   | |||||||  ||||||  ||||||||||||||||    |   |
Db      58490 CTAAATGTGGAATGTTCTGGAGAGCACTATTTTGGGCACTTAATGTCTTCTCCAAAAGGT 58431

Qy        988 CTACTTAGGGCACAGCACT-GAAATTAACACCACTAAGTTTGTCATGTCCATGTAGTTAG 1046
              | |||||||||||||||| ||||||||||  ||| || ||||||  ||||||  ||| |
Db      58430 TTGCTTAGGGCACAGCACTGGAAATTAACACTCCTATGTCTGTCATATCCATGGGGTTGG 58371

Qy       1047 TCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGACCTCTTATGTGTGTCCAGCCTTTCTTC 1106
              |||||  |||   ||||  ||||  ||||  |||  | |||||  |||  ||  |||||||| || |
Db      58370 TCTCATGCAACACAGCTTCAGAAGTGAAACCAATCTCTTCTGTTTGGCCAGCCTTCCTCC 58311

Qy       1107 CTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTCCATAGGAACATCAGT--CCTGAATCC 1164
              |||||||||||||||||| | ||||||||||||||| ||||| | ||||  |  ||  |
Db      58310 CTTCAGAAGTCAGCTGTGGTATCTGCTGACTCTCCACAGGAAAAGCAGTGCCTGGATCCT 58251

Qy       1165 TCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGG 1224
              | |||| || ||||||||||||||  |||||||||||| ||||     |  |  | | |
Db      58250 CAAAACCATCACCTGGAGTAGTAAGTAATCCTGACAGTCCCAGAAAGCATTCAGTGTTCG 58191

Qy       1225 ATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAA 1284
              | || ||   | ||||| | |     ||| ||||||||||||||| ||| ||||| |||
Db      58190 AGCTGCAGGACCTGTGATAGATTACAGGAGTGAAATGAGAAAGCTGAGCTTTTCAGATAA 58131

Qy       1285 ATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAA 1344
              | |||| ||||  |||||  |||| |||||| |||||||||||| ||||| |||||||||
Db      58130 ACCTTGTTTTT--ACAAGTTCCCTCATTTTATTGCATAATTATTCTGAATCCACTGATAA 58073

Qy       1345 TTTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTT 1404
              ||| |||||||||| |||||||||  |||||||| | ||| |||||| ||||||||||
Db      58072 TTTGTACAATTTTCCTATAAGTCATTCACACACAATTCTCTCCCATGCAGCACTTGGCTT 58013

Qy       1405 TGCTAATACATATCTATT-ATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCAC 1463
              ||||||||| ||| |||| |||||||||||| |||| ||| ||||||||||||||| :
Db      58012 TGCTAATACGTATGTATTAATGAGAGCTGTGTTTCTCAAGTGTAAATGTTTTATATGGGT 57953

Qy       1464 TAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCT 1523
              || |||||||||| |||| ||| || ||||||| | | ||||||  |  ||||||||
Db      57952 TATGGCTCTTGGCTGACATGCAAA--GGCTATTGAGTACTCTGATACCAAGATCTTTTCT 57895

Qy       1524 CCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATT 1583
              | |       || || | |||||||||||  |||| ||||| |||||||||| | || |||
Db      57894 CTA-------CACATATCAAGAATTCATGGGATT-GCTGAGATAGACTGATTTATCTATT 57843

Qy       1584 TCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTA 1643
              ||| ||||||| | | | || |   || | ||||  |||| ||||| || |||| ||  |
Db      57842 TCTATGCTCACATGTAACAAAGGCATCCTCAGCTTGGGAAAAAAAAATAAAATCCATACA 57783

Qy       1644 ATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAG 1703
              |||| | |||||  |||  ||| |   ||| ||||||||||| |||  || | ||    |
Db      57782 GTTAAATGAACAGACTGACTTTCCCTGATACAAAAATAAGACAAAAGAAAC-TGCATGAA 57724

Qy       1704 TCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATT---------------------- 1741
                | | ||||| |||||||||||||||||| ||||||
Db      57723 AGACAGTATTTCTTTAATCAGGTCATTGAGAGTCTATTATATATATATATATATATATAT 57664

Qy       1742 ------------------------------------------------------------ 1741

Db      57663 ATATATATATAAAAAATATAATATATATTATATATATTGCATTATATTATATATAATA 57604

Qy       1742 ------------------------------------------------------AAA 1744
                                                                      | |
Db      57603 ATATATGTTATATATCAATTATATATTATAATAATATATTATATATAATATATTATAATA 57544

Qy       1745 AAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTTT 1804
              | |  ||  ||| |||||||||||||||||||| || ||||| ||||| | | || |
Db      57543 TAATTATTATATTATTTATGATGAGAACTATCTTGATTCCAGTGGTCAGTGATAACCTCT 57484

Qy       1805 TTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTACAGGT 1864
              || ||||| ||   ||| ||||| |  ||||||| |||||| ||| || | | |   |
Db      57483 CTGACCTGTTGTATTTACATAGAATGATGGCTTTGCCATAAAATCTCATCTAGCAGGAGC 57424

Qy       1865 GCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTTA 1924
              |  |  |||||||||  ||||||| |||||||||  ||||||| | |||||  |||| |
Db      57423 ACTGAAATAAGTGCCCTGGCTTGCTTTTAAAAGCACTCTGGATTATATTTTAATCTTTCA 57364

Qy       1925 GTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGTGAAG 1984
              |||||   ||| ||||||||||||||||||| |||||   ||| ||||||||||| ||||
Db      57363 GTGAATGAAAGAGAATAAACAAATCCCTCTGTGAGAAAACCTCTTCCATCCTTGGGGAAG 57304

Qy       1985 TCATTCTGCCAGAATTC 2001
              ||||||||||||| |||
Db      57303 TCATTCTGCCAGAGTTC 57287
```

RESULT 10
AC134591

LOCUS	AC134591 222814 bp DNA linear ROD 01-JAN-2004
DEFINITION	Mus musculus BAC clone RP23-376L4 from chromosome 19, complete sequence.
ACCESSION	AC134591
VERSION	AC134591.4 GI:38348187
KEYWORDS	HTG.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 222814)
AUTHORS	VanBrunt,A., Van Brunt,A., Cotton,M., Meyer,R. and Haakenson,W.
TITLE	The sequence of Mus musculus BAC clone RP23-376L4
JOURNAL	Unpublished (2001)
REFERENCE	2 (bases 1 to 222814)
AUTHORS	Wilson,R.
TITLE	Sequencing of Mus musculus
JOURNAL	Unpublished (2001)
REFERENCE	3 (bases 1 to 222814)
AUTHORS	McPherson,J.D. and Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	4 (bases 1 to 222814)
AUTHORS	McPherson,J.D. and Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (28-DEC-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	5 (bases 1 to 222814)
AUTHORS	Wilson,R.K.
TITLE	Direct Submission
JOURNAL	Submitted (15-NOV-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	6 (bases 1 to 222814)
AUTHORS	Wilson,R.
TITLE	Direct Submission
JOURNAL	Submitted (01-JAN-2004) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT	On Nov 15, 2003 this sequence version replaced gi:27413228. ---------- Genome Center Center: Washington University Genome Sequencing Center Center code: WUGSC Web site: http://genome.wustl.edu Contact: submissions@watson.wustl.edu ---------- Summary Statistics Center project name: M_BA0376L04 ----------. NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (http://www.chori.org) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

FEATURES Location/Qualifiers

```
     source          1. .222814
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="19"
                     /map="19"
                     /clone="RP23-376L4"
                     /clone_lib="RPCI-23"
     repeat_region   1. .55
                     /rpt_family="Alu"
     repeat_region   147. .265
                     /rpt_family="L1"
     repeat_region   404. .543
                     /rpt_family="MaLR"
     repeat_region   528. .841
                     /rpt_family="MaLR"
     repeat_region   1425. .1598
                     /rpt_family="B2"
     repeat_region   1635. .1800
                     /rpt_family="L1"
     repeat_region   2314. .2430
                     /rpt_family="Alu"
     repeat_region   2986. .3130
                     /rpt_family="B4"
     repeat_region   4371. .4534
                     /rpt_family="MIR"
     repeat_region   5061. .5126
                     /rpt_family="MER2_type"
     repeat_region   6039. .6273
                     /rpt_family="B4"
     repeat_region   6287. .6631
                     /rpt_family="MaLR"
     repeat_region   6642. .6790
                     /rpt_family="B4"
     repeat_region   7481. .7545
                     /rpt_family="ERV1"
     repeat_region   8637. .8698
                     /rpt_family="Alu"
     repeat_region   8773. .9062
                     /rpt_family="B4"
     repeat_region   9142. .9552
                     /rpt_family="MaLR"
     repeat_region   9553. .9600
                     /rpt_family="ID"
     repeat_region   9601. .9615
                     /rpt_family="B4"
     repeat_region   10198. .10331
                     /rpt_family="Alu"
     repeat_region   10429. .10514
                     /rpt_family="ERVK"
     repeat_region   10470. .10902
                     /rpt_family="ERVK"
     repeat_region   12643. .12771
                     /rpt_family="B4"
     repeat_region   16204. .16251
                     /rpt_family="ERV1"
     repeat_region   20296. .20440
                     /rpt_family="Alu"
     repeat_region   20738. .21229
                     /rpt_family="ERVL"
     repeat_region   26642. .27133
                     /rpt_family="ERVL"
     repeat_region   27205. .27409
                     /rpt_family="B2"
     repeat_region   27445. .27588
                     /rpt_family="B4"
     repeat_region   27764. .27813
                     /rpt_family="ERV1"
     repeat_region   27917. .28095
                     /rpt_family="ERVK"
     repeat_region   28021. .28327
                     /rpt_family="MaLR"
     repeat_region   28540. .28660
                     /rpt_family="MaLR"
     repeat_region   28897. .29037
                     /rpt_family="Alu"
     repeat_region   29352. .29504
                     /rpt_family="Alu"
     repeat_region   29543. .29676
                     /rpt_family="Alu"
     repeat_region   31118. .31199
                     /rpt_family="L2"
     unsure          31652
                     /note="Unresolved bases."
     repeat_region   32320. .32524
                     /rpt_family="B2"
     repeat_region   32542. .32598
                     /rpt_family="Alu"
     repeat_region   34883. .35188
                     /rpt_family="MaLR"
     repeat_region   35500. .35695
                     /rpt_family="MaLR"
     repeat_region   35787. .36037
                     /rpt_family="MaLR"
     repeat_region   36446. .36625
                     /rpt_family="MaLR"
     repeat_region   37694. .38232
                     /rpt_family="L1"
     repeat_region   39171. .39242
                     /rpt_family="Alu"
     repeat_region   39469. .39647
                     /rpt_family="MIR"
     repeat_region   39648. .39764
                     /rpt_family="Alu"
     repeat_region   39765. .39827
                     /rpt_family="MIR"
     repeat_region   39942. .40150
                     /rpt_family="B2"
     repeat_region   40468. .40550
                     /rpt_family="L2"
     repeat_region   40619. .41063
                     /rpt_family="MaLR"
     repeat_region   41084. .41408
                     /rpt_family="MaLR"
     repeat_region   41409. .41757
                     /rpt_family="MaLR"
     repeat_region   41948. .42018
                     /rpt_family="L2"
     repeat_region   42080. .42185
                     /rpt_family="MIR"
     repeat_region   43881. .44239
```

```
  Query Match             14.8%;  Score 296;  DB 9;  Length 222814;
  Best Local Similarity   65.9%;  Pred. No. 8.5e-55;
  Matches  618;  Conservative    0;  Mismatches  270;  Indels   50;  Gaps   11;

Qy       1044 TAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGACCTCTTATGTGTGTCCAGCCTTTC 1103
             || ||    |   | ||| ||||| || ||||   ||| ||| |||||| |||||||  |
Db      82177 TACTCATGTGTGATACAGTCTCAGAAGAGGAAGCCACCACTTCTGTGTGGCCAGCCTCCC 82236

Qy       1104 TTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTCCATAGGAACATCAGT-CCTGAAT 1162
              ||    |  |  |   ||  |||      ||  |  ||  |||||    ||| |||||||
Db      82237 CTCTGAAATGACCCTCTTGGCTTTCTATGTGGATGCCCTCAGGAATGCAAGTACCTGAAT 82296

Qy       1163 CCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCT 1222
              || || ||||||  ||| |||| ||||   | ||| |||| |   |||    |  || |
Db      82297 CCCCAAACCACCTCCTGAAGTACTAAG-CATTCTGTCAGTTCAGCAAGCCCCCACCCCCG 82355

Qy       1223 GGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGT 1282
              ||| | |||   | ||||    ||  |||||||||||||||| | |  ||| ||||| |
Db      82356 GGAGCCCCAGTTCCTGTGGTGAACTATGAGAGAGAAATGAGAAGGGTCAGCTTTTCAGAT 82415

Qy       1283 AAA-------------TCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTATTT 1329
              |||             | |||||||||  |   |  ||| || ||| |||||||| ||  ||
Db      82416 AAACTTTGCCTCTTTGTTTTGCTTTTTAAACAAGCCCATAGTTTCACTGCATAGTTGCTT 82475

Qy       1330 TGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCA 1389
              ||||| ||||||||||||   |||||| ||  ||||||||  ||| | | | |||||| |
Db      82476 TGAATCCACTGATAATTTGGGCAATTTCCCTGTAAGTCATTCACAAATAGTTCCCTCTGA 82535

Qy       1390 TGCAACACTTGGCTTTGCTAATACATATCT-----ATTATGAGAGCTGTGCTTCTTAAGC 1444
              ||   | ||  ||| |||    || |||||       | ||||||||||||||||  ||
Db      82536 TGTTGCCCTCAGCTATGCGGCCACGTATCTGTCTATTACTGAGAGCTGTGCTTCTTGAGT 82595

Qy       1445 GTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATG 1504
              |||||||||||| |||| |||||||||||||||||||||  |||||| ||||  ||| |
Db      82596 ATAAATGTTTTATGTGCATTAAGGCTCTTGGCTTACATAT-CAAGGGGCATTGGACAACG 82654

Qy       1505 TGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAA 1564
              |||||||||| ||||   ||  | | ||| | ||||||||||||| | | |||| ||||
Db      82655 AGATACAGAAGCCTTT--GCCTGACGACTCGTGTGTAAAGAATTCACTGGGTTGGGTGAA 82712

Qy       1565 ATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTA----GCTAAG 1620
              ||| |||||||||||||||  |||||||||      |||||||||| ||     | | |
Db      82713 ATACACTGATCTGTCCATTTCCGTGCTCACTTCCAGCAAGGAAGTCACTATCAGCCAAGG 82772

Qy       1621 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 1680
              | | |||  |||  |||||| ||||| ||||||| | ||| ||||||||||| | ||
Db      82773 AACCGAAAGGTACTGTCTATGCAATTAAAAGAACAGGTTGGCTTTGCTCAATACAGAA-- 82830

Qy       1681 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 1740
                | ||||||||||||  | || || | | ||| ||||||  |||  | |||||  |||||
Db      82831 ---ACAAAGAAACCATGCAAGAGACACAGTGTTTCTTTAATTAGGAGACTGAGATTCTAT 82887

Qy       1741 TAAA-------------AAGTATTTGAATTCTTTATGATGAGAACTAT-CTTGACTCAA 1785
              | ||               |    |||   |||||||| ||  | |||| ||||||||||
Db      82888 TGAAGACACTTTATGATTATTATTTTTTTTTCTTTATGGTGGCAGCTATCCTTGACTCAA 82947

Qy       1786 GTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAA 1845
              ||||  |||| ||  ||  |   | || | |  |||| | ||| | | || ||||  ||||
Db      82948 GTGGGAAGTGTTGGCCTCT---GGCTTTTGCACCTATGCAGAGAAGTGGTTTTGCTATAA 83004

Qy       1846 AGTCTTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGG 1905
              |  || | ||||| | ||| |  |  | || |  | |||||||| ||| |   |    |
Db      83005 ACGCTCAAATGGTGCCGGTTCGGAACTCAGGGTTCCAGCTTGCTCCTAACATTCTGGATG 83064

Qy       1906 ATTTTGTTTTAGACTTTTAGTGAACTGAAGGGAATAAA 1943
              ||||   |        ||       | ||||  | ||||
Db      83065 GTTTTTAATCCAGGAACTACCTGTTTCAAGGCATTAAA 83102


RESULT 11
AC128358
LOCUS       AC128358               237778 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-523J8, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC128358
VERSION     AC128358.3  GI:25007897
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 237778)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 237778)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 237778)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:22856031.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GZCK
                Center clone name: CH230-523J8
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 206828 bases at least Q40
                Consensus quality: 209779 bases at least Q30
                Consensus quality: 211515 bases at least Q20
                Estimated insert size: 210430; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   137365: contig of 137365 bp in length
            *  137366   137465: gap of unknown length
            *  137466   237778: contig of 100313 bp in length.
FEATURES             Location/Qualifiers
     source          1. .237778
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-523J8"
     misc_feature    1. .1519
                     /note="wgs_contig"
     gap             137366. .137465
                     /estimated_length=unknown
     misc_feature    137466. .139824
                     /note="wgs_contig"
     misc_feature    203894. .205175
                     /note="wgs_contig"
ORIGIN

  Query Match             13.2%;  Score 263.4;  DB 14;  Length 237778;
  Best Local Similarity   65.8%;  Pred. No. 1.3e-47;
  Matches  590;  Conservative    0;  Mismatches  251;  Indels   56;  Gaps   12;

Qy       1040 TAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGACCTCTTATGTGTGTCCAGCC 1099
              ||| || ||    |   | ||| ||||| || ||||   ||| ||| |||||| || |||
Db     211255 TAGATACTCATGAGTGATACAGTCTCAGAAGAGGAAGCTACCACTTCTGTGTGGCCTGCC 211314

Qy       1100 TTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTCCATAGGAACATCAGTCCTG 1159
              || | ||  | | |   |   ||  || | ||  ||  | |    | |        ||||
Db     211315 TTCCCTC--TAAAACCACCCATGCTTTCTATGTGGATGCCCTCACGAATGTAAGTACCTG 211372

Qy       1160 AATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACC 1219
              || || || ||||||  ||| ||||||     |  ||| |||  |   ||| | || |||

Db     211373 AAACCCCAAACCACCTCCTGAAGTAGTGGCCACA-CTGTCAGCTCAGCAAGCTCTC-ACC 211430

Qy       1220 GCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCA 1279
              | ||| | || |    ||||    ||  ||||| |||||| ||||   |  ||| ||||
Db     211431 CCAGGACCCCCGATTTCTGTGGTGAACTATGAGAAAGAAATAAGAAGAATCAGCTTTTCA 211490

Qy       1280 GGTAAATCTTGCTTTTT-------------------------CACAAGCCCCCTAATTT 1313
              | ||||  |||| | ||                        ||| |  |||||| |||
Db     211491 GATAAACTTTGCCTCTTTAAAAAAAAAAAACAAAAACAAAAACCACCAAGCCCCTAGTTT 211550

Qy       1314 TACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTAC 1373
              |||| ||| |||||||||||| ||||||||||||   |||||||||  ||||||||  ||
Db     211551 CACTGTATAGTTATTTTGAATCCACTGATAATTTGAGCAATTTTCCTGTAAGTCATTCAC 211610

Qy       1374 ACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCT-----ATTATGAGA 1428
              |      | |||||| ||    ||||  | | ||||   ||||||||       |  |||||
Db     211611 AAGTGGTTCCCTCTAATATCGCACTCAGTTATGCTCCCACATATCTGTCTGTTACTGAGA 211670

Qy       1429 GCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAA 1488
              || |||||||| || |||||||||| || |||| |||||||||| |||||||||  ||
Db     211671 ACTCTGCTTCTTGAGTGTAAATGTTTGATGTGCATTAAGGCTCTTGACTTACATAT-GAA 211729

Qy       1489 GGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATT 1548
              |||| ||||| |||||||| |||||||| ||| |    ||   |||| | | | ||||||
Db     211730 GGGGCATTGAACAATGTGACACAGAAGTATTTGCCTTACGACTCTCGTGTCTGAAGAATC 211789

Qy       1549 CATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAG 1608
              || | | |||| |||||||||| |||||||| |||||| |||||||| | | |||||||
Db     211790 CACTGGCTTGGGTGAAATAGAGTGATCTGTCAATTTCTGTGCTCACTTCTAACAAGGAAG 211849

Qy       1609 TCATTAGCTAAGGA-----ACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTT 1663
              ||| || |    |     | |||||  ||  |||||| ||||| ||||||| | ||| |
Db     211850 TCACTATCAGCCAAGGAACAAAAAAAGGACTGTCTATGCAATTAAAAGAACAGGTTGGCT 211909

Qy       1664 TTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCA 1723
              |||||||||| | ||     | |||||||||||  | || || ||| ||| |||| | |
Db     211910 TTGCTCAATACAGAA-----ACAAAGAAACCATGCAAGAGACACAATGTTTCTTTAGTTA 211964

Qy       1724 GGTCATTGAGAATCTATT--AAAAAGTATTTGAATTCTTTATGATGAGAACTAT-CTTGA 1780
              ||  | |||| ||||||  | | | | |   || |||||| |  || |||| || ||
Db     211965 GGCGACTGAGATTCTATTGAAGACACTTTATTTTTTGTTTATGGTAGGAGCTATCCTGGA 212024

Qy       1781 CTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGT 1840
              ||||||||  |||| |||    ||  | || | | ||||   |||| | |  | ||||
Db     212025 CTCAAGTGGGAAGTGCTGA---CTTCTGGCTTTTGCCCCTGTATAGAGAAGTAGTTTTGC 212081

Qy       1841 CATAAAGTCTTATATGGTACAGGTGCCAAGTTA----AGTGCCCAAGCTTGCTCTTA 1893
              ||||||  || | ||||| |||||     |   |    ||||| | ||||||||||||
Db     212082 CATAAACACTCAAATGGTGCAGGTATGGATGGAACGCAGTGCTCCAGCTTGCTCTTA 212138


RESULT 12
AX033930
LOCUS       AX033930               300 bp    DNA     linear   PAT 21-SEP-2000
DEFINITION  Sequence 22 from Patent WO9851790.
ACCESSION   AX033930
VERSION     AX033930.1  GI:10280499
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Cancilla,M.R., Choo,K.H. and Du,S.D.
  TITLE     A novel nucleic acid molecule
  JOURNAL   Patent: WO 9851790-A 22 19-NOV-1998;
            CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART
            DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN
```

```
  Query Match             4.9%;  Score 97.6;  DB 6;  Length 300;
  Best Local Similarity   66.7%;  Pred. No. 9.7e-11;
  Matches  168;  Conservative   18;  Mismatches   62;  Indels    4;  Gaps    4;

Qy       1753 GAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTG 1812
             ||||||| | ::|| |:|||||||||:||||| | :||   |          |||||||
Db          1 GAATTCTCTGWKATTAKAACTATCTTGMCTCAAATTSACTTGGTGAGCTAACCTGGCCTG 60

Qy       1813 TGGTCCCTACG-TAGAAAGGAGGCTTTGTCATAAAG-TCTTATATGGTACAGGTGCCAAG 1870
             |||||||| | |  || |||||||||||||||| || || |:| |||||| :|||| ||
Db         61 TGGTCCCTTGGCTTTAATGGAGGCTTTGTCATATAGATCATMTGTGGTACTKGTGCCTAG 120

Qy       1871 TT-AAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTTAGTGAA 1929
             ||  |||||||   ||||||:||  : || ||||:|||||:|   || || | :| : ||
Db        121 TTGTAGTGCCCTGCCTTGCTSTTCTWGGCTTACTKGATTTWGGGGTATACATCWATKTAA 180

Qy       1930 CTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGTGAAGTCATT 1989
             :|:|| ||  |   |   |||| ::|||||| |||||||||  ||| || ||| || ||
Db        181 YTSAAAGGTCT-TTCTCCTCCCGYYGGGAGAATTTCTCCTCCTCCCTCGGAGAACTCTTT 239

Qy       1990 CTGCCAGAATTC 2001
             ||:||  |||||
Db        240 CTSCCGAAATTC 251


RESULT 13
I66494
LOCUS       I66494                 7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             3.0%;  Score 60.8;  DB 6;  Length 7218;
  Best Local Similarity   2.9%;  Pred. No. 0.0085;
  Matches   11;  Conservative  226;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        408 CAGGAAAAAGAACATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGC 467
              ||||   | | |  | | || ::: ::  :   : ::::: ::::   :::: : ::: :
Db       1046 CAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1105

Qy        468 TTTTCACAGCTTACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAG 527
              ::::: :  ::: ::: :: ::: : :::    :: : : :::: :     ::::: :
Db       1106 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1165

Qy        528 GGTTTCTACTAAGCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAG 587
                ::::: ::   ::   :::::: : :  ::::::::: ::::: :: :::: ::::
Db       1166 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1225

Qy        588 TGATCTTTCTCTTTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCT 647
              :  :::::::::::::::: :::::    ::  :  :: ::: : :   :: :::  ::
Db       1226 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1285

Qy        648 TTGCTTTGGGTTTTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTAC 707
              :: ::::  :::::::::      :  : :::::   :: ::::   ::   :   :: :
Db       1286 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1345

Qy        708 AGCCACTGGGCCAGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTT 767
              :: ::   ::   ::::::::::::: ::   :: : :::::    :::::: ::::
Db       1346 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1405

Qy        768 AGTTTAGGTCAAGCATTCTT 787
                :::   ::   : :::::
Db       1406 YYYYYYYYYYYYYYYYYYYYY 1425


RESULT 14
AX655393
LOCUS       AX655393                2000 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 5263 from Patent WO03000898.
ACCESSION   AX655393
VERSION     AX655393.1  GI:29158207
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 5263 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

  Query Match             2.9%;  Score 58.8;  DB 6;  Length 2000;
  Best Local Similarity   11.8%;  Pred. No. 0.027;
  Matches   54;  Conservative  215;  Mismatches  187;  Indels    2;  Gaps    1;

Qy       1361 ATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTA 1420
              :::::: :::   :: ::  |   ::::| : |:|::  :|::: |:::    :   :::
Db        508 MYRMRWKRYYMSARYTMRYCARKKYSYSAARKARCWYRGKGYYWAGMWMKRYKRMYMYKM 567

Qy       1421 TTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTAC 1480
              :::::: :  :: |: :::|  :: |:  | :     : :|     ::  : ::|
Db        568 MWWYKRKYSKCSWYCKMSYYASCMKSARKAGAKMCKRSKMSAWSKSMRSSRKCRKCASKR 627

Qy       1481 ATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGT 1540
               |:: |: || |   ::: :   |  :: :| ::   |: :::: :   :: : :::::
Db        628 SSAKRYAMMGGMTSGSRMSRWKSYTCYWRKWGSMKSTCTWMYYMSKYTYAKYGSYWRYRY 687

Qy       1541 AAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCA 1600
              :|: :  : ::: :      : |: | :    :|  ::: :::|:::: |:|  ::: ::
Db        688 RAWCMYMWRWYYYRYRSYMTYMAWYTSSTRMAMTGMKYSGRYWTSWYKYCKCSWKYRSMW 747

Qy       1601 TAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTG 1660
              : :: :: : :: :: :| : :::: :  :::: :   :  :: ::: :::::::  ::
Db        748 YYWSWWWAKTWKMWRRYATRMMWMWYRYSMKWYTWCTMWGYWWYWWRTYMKMRYMW--YK 805

Qy       1661 GTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAA 1720
              |:|: :::| ::|  :||:::    ::  :    |: :|:   ::: :  :   : ::
Db        806 CTKTYWYWSATYWTGTWAAWWMAKTKMRMGMTGAKTRGRARKARYWWKWATWCATKRWMT 865

Qy       1721 TCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGA 1780
              :  | : |:| :   |: : :: :::| : ::: ::| ::| :::|   ::|:::  ::|
Db        866 KGKGAKWAWTWMAKAWRKYYWSWMRAWYYYYKTRRTRYKTCWWKARWGSWAYWRMWWKGS 925

Qy       1781 CTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCC 1818
              :::: ::|  :|: ::   : ||::  |   |:  ||
Db        926 AKMWWMWKGGRWGWTKYWYWYCTTWKMACGRATKYMCC 963


RESULT 15
I66494/c
LOCUS       I66494                 7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
```

```
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             2.9%;  Score 58.4;  DB 6;  Length 7218;
  Best Local Similarity   1.4%;  Pred. No. 0.029;
  Matches    5;  Conservative  216;  Mismatches  127;  Indels    0;  Gaps    0;

Qy          8 TTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCAGGAC 67
              || ||| :::::::::::::::::: :::::: : :  :: : : : : ::   :::: 
Db       1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

Qy         68 CATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATAT 127
               : :::  : ::: :: :::::   ::   :::  :: : :::  :::::::::: : :
Db       1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

Qy        128 TGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAACAAGAGAA 187
              ::   : : ::       : :::::::: :::: ::::::::    ::::: :::::::
Db       1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

Qy        188 TTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCTGTCCCTG 247
                :     :::    ::::: :  ::::  ::   :::: :::  :       :     :
Db       1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

Qy        248 AGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGATGAAGGT 307
              ::: :: :: :: :: :: ::::    : :::::::::::: : : :  :::: :::::
Db       1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

Qy        308 AGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGG 355
              ::::  ::: ::: ::::: :::::  :  : :  : :  ::::: ::
Db       1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1094


RESULT 16
AL137881/c
LOCUS       AL137881              143324 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-40A8 on chromosome 13 Contains a
            novel gene, the 3' end of the GUCY1B2 gene for guanylate cyclase 1
            soluble beta 2 and  a CpG island, complete sequence.
ACCESSION   AL137881
VERSION     AL137881.12  GI:8894206
KEYWORDS    HTG; CpG island; guanylate cyclase; GUCY1B2.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 143324)
  AUTHORS   Pelan,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 1, 2000 this sequence version replaced gi:8546598.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-40A8 is from the library RPCI-11.1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .143324
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-40A8"
                     /clone_lib="RPCI-11.1"
     misc_feature    1
                     /note="Clone_left_end: RP11-40A8"
     misc_feature    14112
                     /note="Clone_right_end: RP11-233H19"
     misc_feature    19526. .19591
                     /note="Single clone region. Region contains reads from
                     clone pcr only (oligos 1&2). Assembly is consistent with
                     restriction digest."
     gene            complement(join(25997. .26264,33228. .33354,54102. .54331))
                     /locus_tag="RP11-40A8.3-001"
     mRNA            complement(join(25997. .26264,33228. .33354,54102. .54331))
                     /locus_tag="RP11-40A8.3-001"
                     /note="match: ESTs: BF574759.1"
     gene            join(53410. .53759,71026. .71097,73094. .73201,74302. .74378,
                     78504. .78618,86940. .87013,89046. .89151,91606. .91687,
                     93082. .93124,97524. .97604,99977. .100383)
                     /locus_tag="RP11-40A8.1-001"
     mRNA            join(53410. .53759,71026. .71097,73094. .73201,74302. .74378,
                     78504. .78618,86940. .87013,89046. .89151,91606. .91687,
                     93082. .93124,97524. .97604,99977. .100383)
                     /locus_tag="RP11-40A8.1-001"
                     /note="match: ESTs: AL520828 AL539901 AL568105 AU119164
                     BG259079 BG259492 BG261287 BG281634 BG393104 BG724123
                     BI460913 BI754776 BM906357 BM918887 BM999529 BQ016444
                     BQ216110
                     match: cDNAs: AK021774 BC001397 BC005088 BC007332
                     BC010174"
     gene            join(53410. .53759,56614. .56748,71024. .71085)
                     /locus_tag="RP11-40A8.1-002"
     mRNA            join(53410. .53759,56614. .56748,71024. .71085)
                     /locus_tag="RP11-40A8.1-002"
                     /note="match: ESTs: BM847956.1"
     gene            join(53422. .53524,56612. .56748,71024. .71097,73094. .73201,
                     74302. .74378,86940. .87002)
                     /locus_tag="RP11-40A8.1-005"
     mRNA            join(53422. .53524,56612. .56748,71024. .71097,73094. .73201,
                     74302. .74378,86940. .87002)
                     /locus_tag="RP11-40A8.1-005"
                     /note="match: ESTs: BI753018.1"
     CDS             join(53696. .53759,71026. .71097,73094. .73201,74302. .74378,
                     78504. .78618,86940. .87013,89046. .89151,91606. .91687,
                     93082. .93124,97524. .97604,99977. .>100090)
                     /locus_tag="RP11-40A8.1-001"
                     /standard_name="OTTHUMP00000018436"
                     /note="match: proteins: Q9D014 Q9HAF7"
                     /codon_start=1
                     /protein_id="CAI13913.1"
```

```
                /db_xref="GI:55957387"
                /db_xref="UniProt/TrEMBL:Q5TBB1"
                /translation="MAAGVDCGDGVGARQHVFLVSEYLKDASKKMKNGLMFVKLVNPC
                SGEGAIYLFNMCLQQLFEVKVFKEKHHSWFINQSVQSGGLLHFATPVDPLFLLLHYLI
                KADKEGKFQPLDQVVVDNVFPNCILLLKLPGLEKLLHHVTEEKGNPEIDNKKYYKYSK
                EKTLKWLEKKVNQTVAALKTNNVNVSSRVQSTAFFSGDQASTDKEEDYIRYAHGLISD
                YIPKELSDDLSKYLKLPEPSASLPNPPSKKIKLSDEPVEAKEDYTKFNTKDLKTEKKN
                SKMTAAQKALAKVDKSGMKSIDTFFGVKNKKKIGKV"
misc_feature    69479
                /locus_tag="RP11-40A8.1-002"
                /note="Clone_left_end: RP11-547C18"
gene            join(71981. .72127,73094. .73201,74302. .74378,78504. .78618,
                86940. .87013,89046. .89151,93080. .93124,97524. .97604)
                /locus_tag="RP11-40A8.1-004"
mRNA            join(71981. .72127,73094. .73201,74302. .74378,78504. .78618,
                86940. .87013,89046. .89151,93080. .93124,97524. .97604)
                /locus_tag="RP11-40A8.1-004"
                /note="match: ESTs: BM559965.1"
gene            join(73733. .73774,74302. .74378,86940. .87013,89046. .89133)
                /locus_tag="RP11-40A8.1-003"
mRNA            join(73733. .73774,74302. .74378,86940. .87013,89046. .89133)
                /locus_tag="RP11-40A8.1-003"
                /note="match: ESTs: BE972426.1"
gene            join(93101. .93124,93690. .93937,97524. .97604,99977. .100121)
                /locus_tag="RP11-40A8.1-006"
mRNA            join(93101. .93124,93690. .93937,97524. .97604,99977. .100121)
                /locus_tag="RP11-40A8.1-006"
                /note="match: ESTs: AW439381.1"
polyA_signal    100360. .100365
                /locus_tag="RP11-40A8.1-001"
polyA_site      100383
                /locus_tag="RP11-40A8.1-001"
gene            join(complement(AL160157.17:66346. .66401),
                complement(AL160157.17:63216. .63289),
                complement(AL160157.17:48615. .48709),
                complement(AL160157.17:34398. .34554),
                complement(AL160157.17:29989. .30088),
                complement(AL160157.17:28603. .28799),
                complement(AL160157.17:27097. .27208),
                complement(AL160157.17:25729. .25832),
                complement(AL160157.17:24495. .24711),
                complement(AL160157.17:20793. .21113),
                complement(AL160157.17:17311. .17459),
                complement(AL160157.17:16576. .16722),
                complement(AL160157.17:10468. .10622),
                complement(AL160157.17:7385. .7536),
                complement(AL160157.17:7095. .7304),
                complement(AL160157.17:4648. .4971),
                complement(138130. .138430))
                /gene="GUCY1B2"
                /locus_tag="RP11-40A8.2-001"
mRNA            join(complement(AL160157.17:66346. .66401),
                complement(AL160157.17:63216. .63289),
                complement(AL160157.17:48615. .48709),
                complement(AL160157.17:34398. .34554),
                complement(AL160157.17:29989. .30088),
                complement(AL160157.17:28603. .28799),
                complement(AL160157.17:27097. .27208),
                complement(AL160157.17:25729. .25832),
                complement(AL160157.17:24495. .24711),
                complement(AL160157.17:20793. .21113),
                complement(AL160157.17:17311. .17459),
                complement(AL160157.17:16576. .16722),
                complement(AL160157.17:10468. .10622),
                complement(AL160157.17:7385. .7536),
                complement(AL160157.17:7095. .7304),
                complement(AL160157.17:4648. .4971),
                complement(138130. .138430))
                /gene="GUCY1B2"
                /locus_tag="RP11-40A8.2-001"
                /product="guanylate cyclase 1, soluble, beta 2"
                /note="match: ESTs: AI247180 AI792818 AI822009 AW300728
                BG181663 BG680126.1
                match: cDNAs: AF038499"
polyA_site      complement(138130)
                /gene="GUCY1B2"
                /locus_tag="RP11-40A8.2-001"
polyA_signal    complement(138150. .138155)
                /gene="GUCY1B2"
                /locus_tag="RP11-40A8.2-001"
misc_feature    143324
                /note="Clone_right_end: RP11-40A8"
ORIGIN

  Query Match             2.8%;  Score 56.4;  DB 8;  Length 143324;
  Best Local Similarity   62.7%;  Pred. No. 0.055;
  Matches  104;  Conservative    0;  Mismatches   61;  Indels    1;  Gaps    1;

Qy         18 AAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCAGGACCATGGATTGC 77
              ||||       ||| ||    | |||  | ||   || ||   |    | |  || ||||||
Db      33346 AAAAGACTTGAGACAGCTCACAATGAAGTACTCTGTATTAGGCAGGGTAATATTGATTGC 33287

Qy         78 AAGTGACAGAAACTCAAT-TCAAACCAACGTAAGTCAAAAGGAAAATATATTGGCTCATG 136
              |||||||||||||| ||||   ||||  ||   |||||  ||||| |  |||||||||| |
Db      33286 AAGTGACAGAAACCCAATCAGAAACTCACTGGAGTCATGAGGAACACTTATTGGCTCAGG 33227

Qy        137 TAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAACAA 182
              ||||||   |   ||  |||||||   | |   | |||||||  ||
Db      33226 TAACCTAACCCTTGAATGGGCAGGGATGTATCTGGTTTGGGAGGAA 33181


RESULT 17
AC021786
LOCUS       AC021786               163554 bp    DNA     linear   HTG 30-MAR-2000
DEFINITION  Homo sapiens clone RP11-21H8, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC021786
VERSION     AC021786.2  GI:7341964
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 163554)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21H8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163554)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2000 this sequence version replaced gi:6721341.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
```

```
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
          ---------- Project Information
            Center project name: L4019
            Center clone name: 21_H_8
          ---------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 158445 bases at least Q40
            Consensus quality: 161065 bases at least Q30
            Consensus quality: 161997 bases at least Q20
            Insert size: 167000; agarose-fp
            Insert size: 162854; sum-of-contigs
            Quality coverage: 5.3 in Q20 bases; agarose-fp
            Quality coverage: 5.4 in Q20 bases; sum-of-contigs
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 8 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1174: contig of 1174 bp in length
          *     1175     1274: gap of 100 bp
          *     1275     4788: contig of 3514 bp in length
          *     4789     4888: gap of 100 bp
          *     4889    14873: contig of 9985 bp in length
          *    14874    14973: gap of 100 bp
          *    14974    32556: contig of 17583 bp in length
          *    32557    32656: gap of 100 bp
          *    32657    52576: contig of 19920 bp in length
          *    52577    52676: gap of 100 bp
          *    52677    73204: contig of 20528 bp in length
          *    73205    73304: gap of 100 bp
          *    73305   109834: contig of 36530 bp in length
          *   109835   109934: gap of 100 bp
          *   109935   163554: contig of 53620 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163554
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-21H8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1174
                     /note="assembly_fragment"
     gap             1175. .1274
                     /estimated_length=100
     misc_feature    1275. .4788
                     /note="assembly_fragment"
     gap             4789. .4888
                     /estimated_length=100
     misc_feature    4889. .14873
                     /note="assembly_fragment"
     gap             14874. .14973
                     /estimated_length=100
     misc_feature    14974. .32556
                     /note="assembly_fragment"
     gap             32557. .32656
                     /estimated_length=100
     misc_feature    32657. .52576
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             52577. .52676
                     /estimated_length=100
     misc_feature    52677. .73204
                     /note="assembly_fragment"
     gap             73205. .73304
                     /estimated_length=100
     misc_feature    73305. .109834
                     /note="assembly_fragment"
     gap             109835. .109934
                     /estimated_length=100
     misc_feature    109935. .163554
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             2.8%;  Score 56.4;  DB 14;  Length 163554;
  Best Local Similarity   62.7%;  Pred. No. 0.054;
  Matches  104;  Conservative    0;  Mismatches   61;  Indels    1;  Gaps    1;

Qy         18 AAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCAGGACCATGGATTGC 77
              ||||       ||| ||   | |||  | ||   || ||   |    | |  || ||||||
Db      47966 AAAAGACTTGAGACAGCTCACAATGAAGTACTCTGTATTAGGCAGGGTAATATTGATTGC 48025

Qy         78 AAGTGACAGAAACTCAAT-TCAAACCAACGTAAGTCAAAAGGAAAATATATTGGCTCATG 136
              |||||||||||||| ||||   ||||  ||   |||||  ||||| |  |||||||||| |
Db      48026 AAGTGACAGAAACCCAATCAGAAACTCACTGGAGTCATGAGGAACACTTATTGGCTCAGG 48085

Qy        137 TAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAACAA 182
              ||||||   |   ||  |||||||   | |   | |||||||  ||
Db      48086 TAACCTAACCCTTGAATGGGCAGGGATGTATCTGGTTTGGGAGGAA 48131


RESULT 18
AC008180
LOCUS       AC008180              193035 bp    DNA     linear   PRI 29-MAY-2002
DEFINITION  Homo sapiens 3 BAC RP11-292L5 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC008180
VERSION     AC008180.15  GI:9558559
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 193035)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Garcia,D.K., Gorrell,J.H.,
            Gorrell,L.L., Guevara,W., Harris,K., He,X., Hernandez,J.,
            Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E.,
            Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y.,
            Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O.,
            Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P.,
            Mei,G., Moore,S., Moorish,T., Morgan,M., Morris,S., Nash,S.,
            Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B.,
            Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D.,
            Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H.,
            Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T.,
            Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
            Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
            Wrensford,G., Yu,W., Zhou,X., Naylor,S.L., Nelson,D. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193035)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 193035)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 193035)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 193035)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 28, 2000 this sequence version replaced gi:9438809.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .193035
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-292L5"
     misc_feature    1. .2005
                     /note="overlaps bases 88684. .90688 of clone AC119732"
                     /function="clone overlap"
     repeat_region   1. .1169
                     /rpt_family="L1"
     repeat_region   complement(1172. .1396)
                     /rpt_family="AluJo"
     repeat_region   complement(1398. .1800)
                     /rpt_family="L1"
     repeat_region   complement(1811. .2094)
                     /rpt_family="L1M2"
     repeat_region   2236. .2775
                     /rpt_family="L1MA4A"
     repeat_region   complement(2820. .3029)
                     /rpt_family="L1MC4"
     repeat_region   3030. .3081
                     /rpt_family="(TAGA)n"
     repeat_region   complement(3082. .3242)
                     /rpt_family="L1MC4"
     repeat_region   complement(3496. .3890)
                     /rpt_family="LTR16A"
     repeat_region   complement(3953. .4079)
                     /rpt_family="MER3"
     repeat_region   complement(5947. .6007)
                     /rpt_family="MER94"
     repeat_region   6130. .6505
                     /rpt_family="MLT1B"
     repeat_region   7376. .7410
                     /rpt_family="(A)n"
     repeat_region   7414. .7467
                     /rpt_family="(CA)n"
     repeat_region   complement(9799. .10540)
                     /rpt_family="L1MB6"
     repeat_region   complement(10543. .10958)
                     /rpt_family="L1M4"
     repeat_region   11299. .11329
                     /rpt_family="(TG)n"
     repeat_region   12452. .12472
                     /rpt_family="(T)n"
     repeat_region   complement(13598. .13880)
                     /rpt_family="MIR"
     repeat_region   13925. .13951
                     /rpt_family="AT_rich"
     repeat_region   13989. .14014
                     /rpt_family="AT_rich"
     repeat_region   14926. .14974
                     /rpt_family="AT_rich"
     repeat_region   15078. .15108
                     /rpt_family="(CA)n"
     repeat_region   complement(15519. .15583)
                     /rpt_family="MIR"
     repeat_region   complement(15702. .15859)
                     /rpt_family="MIR"
     repeat_region   complement(16543. .16826)
                     /rpt_family="L1PA7"
     repeat_region   complement(16843. .17113)
                     /rpt_family="AluSp"
     repeat_region   complement(17203. .17655)
                     /rpt_family="MLT1C"
     repeat_region   20097. .20232
                     /rpt_family="(TA)n"
     repeat_region   complement(20777. .21168)
                     /rpt_family="LTR52"
     repeat_region   21905. .21935
                     /rpt_family="(TG)n"
     repeat_region   22022. .22154
                     /rpt_family="FLAM_A"
     repeat_region   complement(23446. .23611)
                     /rpt_family="FAM"
     repeat_region   23653. .23818
                     /rpt_family="(TA)n"
     repeat_region   23819. .23855
                     /rpt_family="(TG)n"
     repeat_region   23856. .23864
                     /rpt_family="(TA)n"
     repeat_region   complement(23865. .24165)
                     /rpt_family="AluSx"
     repeat_region   24375. .24584
                     /rpt_family="MIR"
     repeat_region   25099. .25372
                     /rpt_family="AluJo"
     STS             25316. .25611
                     /standard_name="G08219"
                     /db_xref="dbSTS:14597"
     repeat_region   25373. .25407
                     /rpt_family="(TAA)n"
     repeat_region   25408. .25422
```

```
                     /rpt_family="AluJo"
     repeat_region   complement(28866. .29170)
                     /rpt_family="HAL1"
     repeat_region   complement(29224. .29489)
                     /rpt_family="HAL1"
     repeat_region   complement(29537. .29606)
                     /rpt_family="L1MD1"
     repeat_region   29607. .29645
                     /rpt_family="(TG)n"
     repeat_region   complement(29646. .29661)

  Query Match              2.7%;  Score 55;  DB 8;  Length 193035;
  Best Local Similarity   63.0%;  Pred. No. 0.11;
  Matches   85;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          4 TTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCA 63
              || ||  | |   ||| |||| ||   |||| | | |||       | ||||
Db     169904 TTAATCAGGAATTGAATAAGAGAACATGAGTCAGATATAATGGCTTGGGCATCAGTCAGG 169963

Qy         64 GGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAAT 123
               | |||| |||| |||||||||| ||  ||| ||||||   | |||   |||| ||||||
Db     169964 AGTCCATAGATTACAAGTGACAGGAAGCCAACTCAAACTTGCTTAAACAAAAATGAAAAT 170023

Qy        124 ATATTGGCTCATGTA 138
               |||||| |||| ||
Db     170024 TTATTGGTTCATATA 170038


RESULT 19
AC130404
LOCUS       AC130404              135653 bp    DNA     linear   PRI 28-FEB-2003
DEFINITION  Homo sapiens 12 BAC RP11-337L12 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC130404
VERSION     AC130404.8  GI:28436243
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 135653)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
            Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
            Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
            Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
            Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
            Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
            King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
            Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
            Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X.,
            Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
            Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
            Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
            Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
            Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
            Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 135653)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 135653)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 135653)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 135653)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 20, 2003 this sequence version replaced gi:28269366.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
```

```
          http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
          ml.
FEATURES             Location/Qualifiers
     source          1. .135653
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-337L12"
     misc_feature    complement(1. .2004)
                     /note="overlaps bases 1. .2004 of clone AC139767"
                     /function="clone overlap"
     repeat_region   96. .234
                     /rpt_family="MIR"
     repeat_region   395. .419
                     /rpt_family="AT_rich"
     repeat_region   complement(721. .1446)
                     /rpt_family="LTR8"
     repeat_region   1460. .2018
                     /rpt_family="L1P5"
     repeat_region   2131. .2164
                     /rpt_family="(TA)n"
     repeat_region   2174. .2472
                     /rpt_family="AluSq"
     repeat_region   2480. .2574
                     /rpt_family="(TA)n"
     repeat_region   2704. .2775
                     /rpt_family="L1MC5"
     repeat_region   complement(3123. .3616)
                     /rpt_family="MER41B"
     repeat_region   complement(3727. .4262)
                     /rpt_family="LTR49"
     repeat_region   complement(4263. .4576)
                     /rpt_family="AluY"
     repeat_region   complement(4577. .4603)
                     /rpt_family="LTR49"
     repeat_region   complement(4955. .5149)
                     /rpt_family="MER4A2"
     repeat_region   5150. .5425
                     /rpt_family="AluSc"
     repeat_region   complement(5426. .5707)
                     /rpt_family="MER4A2"
     repeat_region   complement(5731. .5951)
                     /rpt_family="MER58A"
     repeat_region   complement(5952. .6050)
                     /rpt_family="MER113"
     repeat_region   6051. .6352
                     /rpt_family="AluSg"
     repeat_region   complement(6353. .6457)
                     /rpt_family="MER113"
     repeat_region   complement(7818. .7961)
                     /rpt_family="L2"
     repeat_region   complement(8067. .8585)
                     /rpt_family="MER1A"
     repeat_region   9194. .9224
                     /rpt_family="(TG)n"
     repeat_region   complement(9514. .9676)
                     /rpt_family="L2"
     repeat_region   complement(9867. .10172)
                     /rpt_family="AluSx"
     repeat_region   10639. .10781
                     /rpt_family="(TA)n"
     repeat_region   11252. .11630
                     /rpt_family="MSTA"
     repeat_region   complement(11635. .11929)
                     /rpt_family="MSTA"
     repeat_region   12496. .12534
                     /rpt_family="(CA)n"
     repeat_region   12535. .12574
                     /rpt_family="(CA)n"
     repeat_region   12667. .12895
                     /rpt_family="MER4D"
     repeat_region   12895. .13532
                     /rpt_family="MER4D"
     repeat_region   complement(14783. .15082)
                     /rpt_family="AluJo"
     repeat_region   18064. .18242
                     /rpt_family="MIR"
     repeat_region   complement(18245. .18293)
                     /rpt_family="L1P4"
     repeat_region   18294. .18565
                     /rpt_family="L1PA13"
     repeat_region   complement(19730. .19832)
                     /rpt_family="L2"

  Query Match             2.7%;  Score 53.8;  DB 8;  Length 135653;
  Best Local Similarity   62.0%;  Pred. No. 0.21;
  Matches   85;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          4 TTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCA 63
              ||  ||||| |||  | ||  ||| ||  |  || |||  | | |  || ||   |
Db      49921 TTCTTTTGTCTGATTATAAACAAAGGATTGCTAATAATTCACTGAAATGTATTAGTTGTG 49980

Qy         64 GGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAAT 123
              |     | ||||||||||||||||||| | | ||||||     ||||  ||| ||  |||
Db      49981 GATATGTTAATTGCAAGTGACAGAAACCCTACTCAAACTGGGTTAAGCAAAATGGGGAAT 50040

Qy        124 ATATTGGCTCATGTAAC 140
               |||||| ||||  |||
Db      50041 TTATTGGTTCATAAAAC 50057


RESULT 20
AC023243/c
LOCUS       AC023243              151751 bp    DNA     linear   HTG 01-MAR-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-329P13 map 2, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC023243
VERSION     AC023243.2  GI:7139101
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 151751)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 2, clone RP11-329P13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151751)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 1, 2000 this sequence version replaced gi:6958060.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
```

```
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
        -------------- Project Information
            Center project name: L3747
            Center clone name: 329_P_13
        -------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 141366 bases at least Q40
            Consensus quality: 146249 bases at least Q30
            Consensus quality: 148083 bases at least Q20
            Insert size: 154000; agarose-fp
            Insert size: 149951; sum-of-contigs
            Quality coverage: 4.2 in Q20 bases; agarose-fp
            Quality coverage: 4.3 in Q20 bases; sum-of-contigs
        --------------.
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 19 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1     1186: contig of 1186 bp in length
        *     1187     1286: gap of 100 bp
        *     1287     2716: contig of 1430 bp in length
        *     2717     2816: gap of 100 bp
        *     2817     3885: contig of 1069 bp in length
        *     3886     3985: gap of 100 bp
        *     3986     4428: contig of 443 bp in length
        *     4429     4528: gap of 100 bp
        *     4529     6662: contig of 2134 bp in length
        *     6663     6762: gap of 100 bp
        *     6763     9149: contig of 2387 bp in length
        *     9150     9249: gap of 100 bp
        *     9250    12293: contig of 3044 bp in length
        *    12294    12393: gap of 100 bp
        *    12394    16919: contig of 4526 bp in length
        *    16920    17019: gap of 100 bp
        *    17020    22858: contig of 5839 bp in length
        *    22859    22958: gap of 100 bp
        *    22959    28829: contig of 5871 bp in length
        *    28830    28929: gap of 100 bp
        *    28930    38397: contig of 9468 bp in length
        *    38398    38497: gap of 100 bp
        *    38498    46298: contig of 7801 bp in length
        *    46299    46398: gap of 100 bp
        *    46399    55275: contig of 8877 bp in length
        *    55276    55375: gap of 100 bp
        *    55376    65891: contig of 10516 bp in length
        *    65892    65991: gap of 100 bp
        *    65992    78122: contig of 12131 bp in length
        *    78123    78222: gap of 100 bp
        *    78223    92076: contig of 13854 bp in length
        *    92077    92176: gap of 100 bp
        *    92177   109888: contig of 17712 bp in length
        *   109889   109988: gap of 100 bp
        *   109989   128244: contig of 18256 bp in length
        *   128245   128344: gap of 100 bp
        *   128345   151751: contig of 23407 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151751
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-329P13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1186
                     /note="assembly_fragment"
     gap             1187. .1286
                     /estimated_length=100
     misc_feature    1287. .2716
                     /note="assembly_fragment"
     gap             2717. .2816
                     /estimated_length=100
     misc_feature    2817. .3885
                     /note="assembly_fragment"
     gap             3886. .3985
                     /estimated_length=100
     misc_feature    3986. .4428
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     gap             4429. .4528
                     /estimated_length=100
     misc_feature    4529. .6662
                     /note="assembly_fragment"
     gap             6663. .6762
                     /estimated_length=100
     misc_feature    6763. .9149
                     /note="assembly_fragment"
     gap             9150. .9249
                     /estimated_length=100
     misc_feature    9250. .12293
                     /note="assembly_fragment"
     gap             12294. .12393
                     /estimated_length=100
     misc_feature    12394. .16919
                     /note="assembly_fragment"
     gap             16920. .17019
                     /estimated_length=100
     misc_feature    17020. .22858
                     /note="assembly_fragment"
     gap             22859. .22958
                     /estimated_length=100
     misc_feature    22959. .28829
                     /note="assembly_fragment"
     gap             28830. .28929
                     /estimated_length=100
     misc_feature    28930. .38397
                     /note="assembly_fragment"
     gap             38398. .38497
                     /estimated_length=100
     misc_feature    38498. .46298
                     /note="assembly_fragment"
     gap             46299. .46398
                     /estimated_length=100
     misc_feature    46399. .55275
                     /note="assembly_fragment"
     gap             55276. .55375
                     /estimated_length=100
     misc_feature    55376. .65891
                     /note="assembly_fragment"
     gap             65892. .65991
                     /estimated_length=100
     misc_feature    65992. .78122
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             78123. .78222
                     /estimated_length=100
     misc_feature    78223. .92076
                     /note="assembly_fragment"
     gap             92077. .92176
                     /estimated_length=100
     misc_feature    92177. .109888
                     /note="assembly_fragment"
     gap             109889. .109988
                     /estimated_length=100
     misc_feature    109989. .128244
                     /note="assembly_fragment"
```

```
     gap             128245. .128344
                     /estimated_length=100
     misc_feature    128345. .151751
                     /note="assembly_fragment"
ORIGIN

  Query Match             2.7%;  Score 53.8;  DB 14;  Length 151751;
  Best Local Similarity   62.0%;  Pred. No. 0.2;
  Matches   85;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          4 TTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCA 63
              ||  ||||| |||  | ||  ||| ||  |  || |||  | | |  || ||    |
Db       6443 TTCTTTTGTCTGATTATAAACAAAGGATTGCTAATAATTCACTGAAATGTATTAGTTGTG 6384

Qy         64 GGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAAT 123
              |     |  |||||||||||||||||| | | ||||||     ||||  ||| ||  |||
Db       6383 GATATGTTAATTGCAAGTGACAGAAACCCTACTCAAACTGGGTTAAGCAAAATGGGGAAT 6324

Qy        124 ATATTGGCTCATGTAAC 140
               |||||| ||||  |||
Db       6323 TTATTGGTTCATAAAAC 6307


RESULT 21
AC041019/c
LOCUS       AC041019              158059 bp    DNA     linear   HTG 22-MAY-2000
DEFINITION  Homo sapiens clone RP11-572N16, WORKING DRAFT SEQUENCE, 11
            unordered pieces.
ACCESSION   AC041019
VERSION     AC041019.2  GI:8016476
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 158059)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-572N16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158059)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 22, 2000 this sequence version replaced gi:7534221.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9578
                Center clone name: 572_N_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 150828 bases at least Q40
                Consensus quality: 154584 bases at least Q30
                Consensus quality: 156053 bases at least Q20
                Insert size: 163000; agarose-fp
                Insert size: 157059; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases; agarose-fp
                Quality coverage: 5.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3521: contig of 3521 bp in length
            *     3522     3621: gap of 100 bp
            *     3622     8787: contig of 5166 bp in length
            *     8788     8887: gap of 100 bp
            *     8888    15129: contig of 6242 bp in length
            *    15130    15229: gap of 100 bp
            *    15230    23646: contig of 8417 bp in length
            *    23647    23746: gap of 100 bp
            *    23747    34488: contig of 10742 bp in length
            *    34489    34588: gap of 100 bp
            *    34589    45971: contig of 11383 bp in length
            *    45972    46071: gap of 100 bp
            *    46072    61548: contig of 15477 bp in length
            *    61549    61648: gap of 100 bp
            *    61649    76890: contig of 15242 bp in length
            *    76891    76990: gap of 100 bp
            *    76991    93810: contig of 16820 bp in length
            *    93811    93910: gap of 100 bp
            *    93911   111867: contig of 17957 bp in length
            *   111868   111967: gap of 100 bp
            *   111968   158059: contig of 46092 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158059
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-572N16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .3521
                     /note="assembly_fragment"
     gap             3522. .3621
                     /estimated_length=100
     misc_feature    3622. .8787
                     /note="assembly_fragment"
     gap             8788. .8887
                     /estimated_length=100
     misc_feature    8888. .15129
                     /note="assembly_fragment"
     gap             15130. .15229
                     /estimated_length=100
     misc_feature    15230. .23646
                     /note="assembly_fragment"
     gap             23647. .23746
                     /estimated_length=100
     misc_feature    23747. .34488
                     /note="assembly_fragment"
     gap             34489. .34588
                     /estimated_length=100
     misc_feature    34589. .45971
```

```
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     gap             45972. .46071
                     /estimated_length=100
     misc_feature    46072. .61548
                     /note="assembly_fragment"
     gap             61549. .61648
                     /estimated_length=100
     misc_feature    61649. .76890
                     /note="assembly_fragment"
     gap             76891. .76990
                     /estimated_length=100
     misc_feature    76991. .93810
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     gap             93811. .93910
                     /estimated_length=100
     misc_feature    93911. .111867
                     /note="assembly_fragment"
     gap             111868. .111967
                     /estimated_length=100
     misc_feature    111968. .158059
                     /note="assembly_fragment"
ORIGIN

  Query Match              2.7%;  Score 53.8;  DB 14;  Length 158059;
  Best Local Similarity   62.0%;  Pred. No. 0.2;
  Matches   85;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          4 TTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCA 63
              ||  ||||| |||  | ||  ||| ||  |  || |||  | | |  || ||    |
Db     113993 TTCTTTTGTCTGATTATAAACAAAGGATTGCTAATAATTCACTGAAATGTATTAGTTGTG 113934

Qy         64 GGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAAT 123
              |      |  |||||||||||||||||| | | ||||||     ||||  ||| ||  |||
Db     113933 GATATGTTAATTGCAAGTGACAGAAACCCTACTCAAACTGGGTTAAGCAAAATGGGGAAT 113874

Qy        124 ATATTGGCTCATGTAAC 140
              ||||||  ||||  |||
Db     113873 TTATTGGTTCATAAAAC 113857


RESULT 22
AC024980
LOCUS       AC024980              187341 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-337L12, WORKING DRAFT SEQUENCE, 31
            unordered pieces.
ACCESSION   AC024980
VERSION     AC024980.3  GI:8076935
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 187341)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-337L12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 187341)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7283232.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7062
                Center clone name: 337_L_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 174400 bases at least Q40
                Consensus quality: 179866 bases at least Q30
                Consensus quality: 182287 bases at least Q20
                Insert size: 185000; agarose-fp
                Insert size: 184341; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1003: contig of 1003 bp in length
            *     1004     1103: gap of 100 bp
            *     1104     2892: contig of 1789 bp in length
            *     2893     2992: gap of 100 bp
            *     2993     5079: contig of 2087 bp in length
            *     5080     5179: gap of 100 bp
            *     5180     6911: contig of 1732 bp in length
            *     6912     7011: gap of 100 bp
            *     7012     8719: contig of 1708 bp in length
            *     8720     8819: gap of 100 bp
            *     8820    10691: contig of 1872 bp in length
            *    10692    10791: gap of 100 bp
            *    10792    12292: contig of 1501 bp in length
            *    12293    12392: gap of 100 bp
            *    12393    14456: contig of 2064 bp in length
            *    14457    14556: gap of 100 bp
            *    14557    17814: contig of 3258 bp in length
            *    17815    17914: gap of 100 bp
            *    17915    21497: contig of 3583 bp in length
            *    21498    21597: gap of 100 bp
            *    21598    25074: contig of 3477 bp in length
            *    25075    25174: gap of 100 bp
            *    25175    28807: contig of 3633 bp in length
            *    28808    28907: gap of 100 bp
            *    28908    32610: contig of 3703 bp in length
            *    32611    32710: gap of 100 bp
            *    32711    36954: contig of 4244 bp in length
            *    36955    37054: gap of 100 bp
```

```
            *    37055    41312: contig of 4258 bp in length
            *    41313    41412: gap of 100 bp
            *    41413    44561: contig of 3149 bp in length
            *    44562    44661: gap of 100 bp
            *    44662    49191: contig of 4530 bp in length
            *    49192    49291: gap of 100 bp
            *    49292    52928: contig of 3637 bp in length
            *    52929    53028: gap of 100 bp
            *    53029    56291: contig of 3263 bp in length
            *    56292    56391: gap of 100 bp
            *    56392    63001: contig of 6610 bp in length
            *    63002    63101: gap of 100 bp
            *    63102    69593: contig of 6492 bp in length
            *    69594    69693: gap of 100 bp
            *    69694    75000: contig of 5307 bp in length
            *    75001    75100: gap of 100 bp
            *    75101    83143: contig of 8043 bp in length
            *    83144    83243: gap of 100 bp
            *    83244    90745: contig of 7502 bp in length
            *    90746    90845: gap of 100 bp
            *    90846    98863: contig of 8018 bp in length
            *    98864    98963: gap of 100 bp
            *    98964   108066: contig of 9103 bp in length
            *   108067   108166: gap of 100 bp
            *   108167   118935: contig of 10769 bp in length
            *   118936   119035: gap of 100 bp
            *   119036   130639: contig of 11604 bp in length
            *   130640   130739: gap of 100 bp
            *   130740   144392: contig of 13653 bp in length
            *   144393   144492: gap of 100 bp
            *   144493   159251: contig of 14759 bp in length
            *   159252   159351: gap of 100 bp
            *   159352   187341: contig of 27990 bp in length.
FEATURES             Location/Qualifiers
     source          1. .187341
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-337L12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1003
                     /note="assembly_fragment"
     gap             1004. .1103
                     /estimated_length=100
     misc_feature    1104. .2892
                     /note="assembly_fragment"
     gap             2893. .2992
                     /estimated_length=100
     misc_feature    2993. .5079
                     /note="assembly_fragment"
     gap             5080. .5179
                     /estimated_length=100
     misc_feature    5180. .6911
                     /note="assembly_fragment"
     gap             6912. .7011
                     /estimated_length=100
     misc_feature    7012. .8719
                     /note="assembly_fragment"
     gap             8720. .8819
                     /estimated_length=100
     misc_feature    8820. .10691
                     /note="assembly_fragment"
     gap             10692. .10791
                     /estimated_length=100
     misc_feature    10792. .12292
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     gap             12293. .12392
                     /estimated_length=100
     misc_feature    12393. .14456
                     /note="assembly_fragment"
     gap             14457. .14556
                     /estimated_length=100
     misc_feature    14557. .17814
                     /note="assembly_fragment"
     gap             17815. .17914
                     /estimated_length=100
     misc_feature    17915. .21497
                     /note="assembly_fragment"
     gap             21498. .21597
                     /estimated_length=100
     misc_feature    21598. .25074
                     /note="assembly_fragment"
     gap             25075. .25174
                     /estimated_length=100
     misc_feature    25175. .28807
                     /note="assembly_fragment"
     gap             28808. .28907
                     /estimated_length=100
     misc_feature    28908. .32610
                     /note="assembly_fragment"
     gap             32611. .32710
                     /estimated_length=100
     misc_feature    32711. .36954
                     /note="assembly_fragment"
     gap             36955. .37054
                     /estimated_length=100
     misc_feature    37055. .41312
                     /note="assembly_fragment"

  Query Match             2.7%;  Score 53.8;  DB 14;  Length 187341;
  Best Local Similarity   62.0%;  Pred. No. 0.2;
  Matches   85;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          4 TTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCA 63
              ||  ||||| |||  | ||  ||| ||  |  || |||  | | |  || ||    |
Db      18999 TTCTTTTGTCTGATTATAAACAAAGGATTGCTAATAATTCACTGAAATGTATTAGTTGTG 19058

Qy         64 GGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAAT 123
              |     |  ||||||||||||||||| | | ||||||     ||||  ||| ||  |||
Db      19059 GATATGTTAATTGCAAGTGACAGAAACCCTACTCAAACTGGGTTAAGCAAAATGGGGAAT 19118

Qy        124 ATATTGGCTCATGTAAC 140
               |||||| ||||  |||
Db      19119 TTATTGGTTCATAAAAC 19135


RESULT 23
AC023225
LOCUS       AC023225              190310 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-440M16 map 1, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC023225
VERSION     AC023225.3  GI:8076879
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 190310)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 1, clone RP11-440M16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190310)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
```

```
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7342073.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6684
                Center clone name: 440_M_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 181728 bases at least Q40
                Consensus quality: 185239 bases at least Q30
                Consensus quality: 186925 bases at least Q20
                Insert size: 175000; agarose-fp
                Insert size: 188610; sum-of-contigs
                Quality coverage: 8.5 in Q20 bases; agarose-fp
                Quality coverage: 7.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1095: contig of 1095 bp in length
            *     1096     1195: gap of 100 bp
            *     1196     2767: contig of 1572 bp in length
            *     2768     2867: gap of 100 bp
            *     2868     4870: contig of 2003 bp in length
            *     4871     4970: gap of 100 bp
            *     4971     6216: contig of 1246 bp in length
            *     6217     6316: gap of 100 bp
            *     6317     7525: contig of 1209 bp in length
            *     7526     7625: gap of 100 bp
            *     7626     9452: contig of 1827 bp in length
            *     9453     9552: gap of 100 bp
            *     9553    11875: contig of 2323 bp in length
            *    11876    11975: gap of 100 bp
            *    11976    14007: contig of 2032 bp in length
            *    14008    14107: gap of 100 bp
            *    14108    20618: contig of 6511 bp in length
            *    20619    20718: gap of 100 bp
            *    20719    29063: contig of 8345 bp in length
            *    29064    29163: gap of 100 bp
            *    29164    36474: contig of 7311 bp in length
            *    36475    36574: gap of 100 bp
            *    36575    46685: contig of 10111 bp in length
            *    46686    46785: gap of 100 bp
            *    46786    58540: contig of 11755 bp in length
            *    58541    58640: gap of 100 bp
            *    58641    73486: contig of 14846 bp in length
            *    73487    73586: gap of 100 bp
            *    73587    89995: contig of 16409 bp in length
            *    89996    90095: gap of 100 bp
            *    90096   111543: contig of 21448 bp in length
            *   111544   111643: gap of 100 bp
            *   111644   143863: contig of 32220 bp in length
            *   143864   143963: gap of 100 bp
            *   143964   190310: contig of 46347 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190310
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1"
                     /clone="RP11-440M16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1095
                     /note="assembly_fragment"
     gap             1096. .1195
                     /estimated_length=100
     misc_feature    1196. .2767
                     /note="assembly_fragment"
     gap             2768. .2867
                     /estimated_length=100
     misc_feature    2868. .4870
                     /note="assembly_fragment"
     gap             4871. .4970
                     /estimated_length=100
     misc_feature    4971. .6216
                     /note="assembly_fragment"
     gap             6217. .6316
                     /estimated_length=100
     misc_feature    6317. .7525
                     /note="assembly_fragment"
     gap             7526. .7625
                     /estimated_length=100
     misc_feature    7626. .9452
                     /note="assembly_fragment"
     gap             9453. .9552
                     /estimated_length=100
     misc_feature    9553. .11875
                     /note="assembly_fragment"
     gap             11876. .11975
                     /estimated_length=100
     misc_feature    11976. .14007
                     /note="assembly_fragment"
     gap             14008. .14107
                     /estimated_length=100
     misc_feature    14108. .20618
                     /note="assembly_fragment"
     gap             20619. .20718
                     /estimated_length=100
     misc_feature    20719. .29063
                     /note="assembly_fragment"
     gap             29064. .29163
                     /estimated_length=100
     misc_feature    29164. .36474
                     /note="assembly_fragment"
     gap             36475. .36574
                     /estimated_length=100
     misc_feature    36575. .46685
                     /note="assembly_fragment"
     gap             46686. .46785
                     /estimated_length=100
     misc_feature    46786. .58540
                     /note="assembly_fragment"
     gap             58541. .58640
                     /estimated_length=100
     misc_feature    58641. .73486
                     /note="assembly_fragment"
     gap             73487. .73586
                     /estimated_length=100
     misc_feature    73587. .89995
                     /note="assembly_fragment"
     gap             89996. .90095
                     /estimated_length=100
```

```
     misc_feature    90096. .111543
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     gap             111544. .111643
                     /estimated_length=100
     misc_feature    111644. .143863
                     /note="assembly_fragment"
     gap             143864. .143963
                     /estimated_length=100
     misc_feature    143964. .190310
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
ORIGIN

  Query Match             2.7%;  Score 53.8;  DB 14;  Length 190310;
  Best Local Similarity   62.0%;  Pred. No. 0.2;
  Matches   85;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          4 TTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCA 63
              ||  ||||| |||  | ||  ||| ||  |  || |||  | | |  || ||    |
Db        548 TTCTTTTGTCTGATTATAAACAAAGGATTGCTAATAATTCACTGAAATGTATTAGTTGTG 607

Qy         64 GGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAAT 123
              |      |  |||||||||||||||||| | | ||||||     ||||  ||| ||  |||
Db        608 GATATGTTAATTGCAAGTGACAGAAACCCTACTCAAACTGGGTTAAGCAAAATGGGGAAT 667

Qy        124 ATATTGGCTCATGTAAC 140
              |||||| ||||  |||
Db        668 TTATTGGTTCATAAAAC 684


RESULT 24
AE014826
LOCUS       AE014826               250663 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 14 section 11 of 13 of the
            complete sequence.
ACCESSION   AE014826 AE014187
VERSION     AE014826.1  GI:23497662
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 250663)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 250663)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .250663
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="14"
     repeat_region   complement(99. .160)
                     /rpt_type=tandem
                     /rpt_unit="(taaaaa)n"
     repeat_region   complement(191. .258)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(273. .394)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(350. .468)
                     /rpt_type=tandem
                     /rpt_unit="(tataa)n"
     repeat_region   complement(472. .530)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(576. .655)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(679. .752)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   679. .742
                     /rpt_type=tandem
     repeat_region   754. .793
                     /rpt_type=tandem
     repeat_region   complement(754. .792)
                     /rpt_type=tandem
                     /rpt_unit="(taaaa)n"
     repeat_region   complement(814. .945)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   816. .845
                     /rpt_type=tandem
     gene            <869. .>4141
                     /locus_tag="PF14_0589"
     mRNA            <869. .>4141
                     /locus_tag="PF14_0589"
     CDS             869. .4141
                     /locus_tag="PF14_0589"
                     /codon_start=1
                     /product="valine - tRNA ligase, putative"
                     /protein_id="AAN37202.1"
                     /db_xref="GI:23497663"
                     /translation="MNIRLIKKIGINYINKSFNTYCNKKLSKMDMKNIKENIKTMKDA
                     YDPKEVESKWYSFWEENDYFKPKKELIEENKNNEHIKKFVIILPPPNVTGTLHIGHTL
                     TVAIQDSLVRYKRMKNYLTLYIPGTDHAGIATQTVVEKMLYKKENKIRQDYGREEFVK
                     KIYEWKDLHGNKINNQMKRIGASVDWSREYFTMNENLSNAVKEAFIKFYESGLIYRDN
                     RLVAWCPHLKTALSDIEVNLEEIKKPTKIKIPSFDHLVEVGVLYKFFYQIKDSEEKIE
                     IATTRIETMLGDVAVAVHPKDKRYAHLIGKEIVHPFIPNRKIIIIADDFVDMQYGTGA
                     VKITPAHDKNDYDMMKRHNLNYINIFTLDGHINENGGKLFQGQHRFECRFKIQEELKK
                     LNLLSDKVPNPMSLPLCSRTNDIIEYMLIPQWYMNCSELAHQAIQNVKLKELEIIPPQ
                     HVNTWYYWLENVRDWCISRQLWWGHRIPAYKIVFKNDVNHNTDDINHNDDNNDDNNNN
                     NNNNNNLEEQNELWVVGRSYEEGLEKAKSLVGDNKPFELIQDEDVLDTWFSSSLVPFS
                     SLGWPNQTEDLETFFPNTILETGQDILFFWVARMVMVSLHLMKKLPFKTIYLHAMIRD
                     SRGEKMSKSKGNVVDPLDIIDGISLNKLHEKLYEGNLPEKEIKRAIELQKKEFPKGIP
                     ECGTDALRFGLLTYLKQGRNVNLDINRIIGYRHFCNKLWNAVKFFLKTLPDNYDNYNI
                     LIDKPEYVEKLQWEDKWILHKLNVYIKNANDNFESYNFSEVAFASYNFWLYDLCDIYL
                     ELIKPRLNTDSHDNFVSGLKEKSDMNNIEKREVVEQNDNNIESVNTNHINEINKCDNY
                     LDGYGANKTLHVCLDYGLRLLHPISPFITEELYQKISSEPYKFNSISLANYPEYINIW
                     HNENINSEMNKFMNIVKQFRSFISTLEIPPKTKLNCFIAAKHEEDKIFIEKVRTKIEV
                     LAKLSSLSVINYNVEDLSDDLKLQVKKCLKDIVSNQFIIYVQSSEEYLKPLLSSMLNK
                     NKKLQASLDSYLKKINDPNYEQKVPEQVRTMYSEKVEELNSQILSISNIICEINECLK
                     NA"
     repeat_region   complement(1923. .1949)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2320. .2410
                     /rpt_type=tandem
                     /rpt_unit="(taa)n"
     repeat_region   2367. .2403
                     /rpt_type=tandem
     repeat_region   complement(3950. .3970)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(4187. .4256)
                     /note="AT_rich"
```

```
                     /rpt_type=tandem
     repeat_region   complement(4287. .4349)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   4323. .4349
                     /rpt_type=tandem
     repeat_region   complement(4391. .4413)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   4419. .4449
                     /rpt_type=tandem
     repeat_region   complement(4419. .4445)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(4480. .4562)
                     /note="AT_rich"
                     /rpt_type=tandem
     gene            complement(<4588. .>4953)
                     /locus_tag="PF14_0590"
     mRNA            complement(<4588. .>4953)
                     /locus_tag="PF14_0590"
     CDS             complement(4588. .4953)
                     /locus_tag="PF14_0590"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN37203.1"
                     /db_xref="GI:23497664"
                     /translation="MKRTDDKIYVDINNEEEYKNLFDDKNDILYIIDIYTRWCGPCIF
                     TFEMINKIYKNNLIFSENVKLFSICAQTVSSLKNYDNNCKPFYIILKNGEIIQQIQGC
                     NIPLIFSFIDEHLMNKKIN"
     repeat_region   complement(4781. .4809)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(4952. .5000)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(5038. .5093)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   5064. .5093
                     /rpt_type=tandem
     repeat_region   complement(5094. .5138)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   5203. .5243
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(5276. .5307)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   5308. .5334
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     gene            complement(<5401. .>9405)
                     /locus_tag="PF14_0591"
     mRNA            complement(<5401. .>9405)
                     /locus_tag="PF14_0591"
     CDS             complement(5401. .9405)
                     /locus_tag="PF14_0591"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN37204.1"
                     /db_xref="GI:23497665"
                     /translation="MKVDIDGLFKSYDHNETIKYIDKLRNEIIEKEDDIKEFLKKESN
                     NVIDNTIRIKKIYGCVEKTKKNLEDIISNYNKLVDNVKSTQVCNIQDIYKNELNEIEK
                     DIVKNVWLKKYDKYEFLFENPFFEKKNKDGKEGLGNIINNIIINEDASLNTKEESVII
                     KYDYLINKKEKKKKDVIVNYDCFYLNNYINNIYFDINKKIEEENYIYCINKIYRDIFI
                     SLIVLKSILKNEKIQNIYINIFKKPNCNYYLHNFVNKYKHHFSNVILKLIHYSFYNLT
                     YKHSCNFQTIIKSIMILLIFYNQKNTDIINFIHKNKENIKQHDINKNQMNNLFHAILK
                     ARLKLIHNFINKKKNKKCNKGNELDRNSTIQKNELIHNNEYLAFNKLHIYFKKIIYVL
                     LNTKYILLLLIDNSCCPLLHEHIKKKNKDEHIYEKNEICEKQFVYNQIHTNYKIHINK
                     ENQIDVVESYENNFLQYMKNEIFILQDIYNNKEHIKNGIEKNKEYFNYHINNIEDIKE
                     KILNTNTVENFILLYKHYNQCIKNIYKYVIYLLHNYYYTNQYNIISIDFLFNEYTNIK
                     LLYNEIKSTILLPETYFKPFIKPLQHSSIINHNCLININEQLFNYFLFLFFNKSISII
                     PLFDNKPIDQWKNVLIKFFKSFSYHVTTFLTYDNNSYFFINDKIFFSFLLNAFYFYYE
                     FYNIDLEIIYSEFSKIPKHFNNTKYLKDKKKPHNRLSEHVLNNNKNCTQEKLFNNAFK
                     NIIKEKIHPLLFSAYKLILFIEVKKMNTFVNVDFLYMNKNKNLSNFQELYDIFINILN
                     NYDNIKTNTHVNDDVQGYDVLLNLLKGNDNDIYMNININTNNNDIRKSADDVNKNGDD
                     MNKNSDDMNKNSDDLNKNGDDMNKNSDGINNNGDDLNKNGDGINNNGDDMNKNSDGIQ
                     CVHVNYCNDFNDEEYKEFYFYVNLLRYNNKNKFNHFLDGLKNIQIYGESNDMSNIYNN
                     NENVHIINNNNVFNSEPNFCLKDIKCIFFYIVYNIFKMCMKCFCSEYFDNLKCHLYYY
                     LHMLLNYNMEHILCDKINSHLLNIFIFSNIFIIYIEKVMTTEQYKNIIIFIIKTILQK
                     KIYKIYITFIQKVKDIYIIEMEKAQNKNLFNKKITELYHILLFDLYFCEQVLDKNSIT
                     NKSFYEDLIYLYKINEMYYIKSTLYFIDIYKQTYNVYNENKNNNNKNNNNNNNNNISK
                     EEHVQKQNIHPLIHKLYKSYNNFVHNQNEKNNNIMKNISFKFLIDDILVELNKLDNLN
                     NIIYQKHIRSRVHNVIKNTYLLYYLFVQGTTLNNIINTKIEEEQNNLDIFHYNNVLKI
                     DDQSIEEKLFTFFS"
     repeat_region   complement(5542. .5574)
                     /note="AT_rich"
```

```
  Query Match             2.6%;  Score 52.8;  DB 2;  Length 250663;
  Best Local Similarity   52.2%;  Pred. No. 0.32;
  Matches  117;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy       1588 TGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTA 1647
              || | | |  |||| |  |  | || |   ||  || |||||  | ||   | | ||| |
Db     230081 TGTTAAGTATTCATTAAAATATAATAATAAAAAAAAAAAAAAAAAAAAAAACAAGAAATAA 230140

Qy       1648 GAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAA 1707
                || ||| |   | |  | |||| ||  |  || |||| |||| | |  ||| | ||
Db     230141 ATGTAATAAGATATATATATTAAATAAAAGGAAGAGGAAAATAAACAAAGAAAAATTTAA 230200

Qy       1708 AATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATG 1767
                | | ||  | |    |||  |     |  ||||||| | |   |||| |  | |
Db     230201 TCAACGTATTATAGCTAAACATATATGTAGTGATAAAAAGAAATGTTATTCATCCTTACC 230260

Qy       1768 AGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCT 1811
              |  ||  |||  | | ||    |||| |    ||| |||  ||
Db     230261 ATCACATTCTATAGTAAACACAACAGAGAGTGACTTGTTGCACT 230304
```

```
RESULT 25
AL954840
LOCUS       AL954840              167004 bp    DNA     linear   VRT 11-OCT-2003
DEFINITION  Zebrafish DNA sequence from clone DKEY-29K19, complete sequence.
ACCESSION   AL954840
VERSION     AL954840.16  GI:37652324
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 167004)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 13, 2003 this sequence version replaced gi:37650875.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
```

```
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-29K19
            is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1. .167004
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-29K19"
                     /clone_lib="DanioKey"
ORIGIN

  Query Match              2.6%;  Score 52.2;  DB 5;  Length 167004;
  Best Local Similarity   50.6%;  Pred. No. 0.46;
  Matches  126;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy       1547 TTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGA 1606
             ||| |||| |  | |||||| | | | ||        ||      |  | | | |||
Db     112615 TTCTGTAGACTATCGGAAATACAATTAGCTTAAAGGGGCTAGTAACTTTGACCTTAAATT 112674

Qy       1607 AGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTG 1666
              ||  |||   || | | ||   ||   || |  |     | || |  || | |  |||
Db     112675 GGTGTTTAAAAAATGTAAAACTGCTTTTATTTTAGCCGAAATAAAACAAATCAGACTTTT 112734

Qy       1667 CTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGT 1726
                |  |||||||  |  |  ||| |   ||||||| |      |  |   |||| ||
Db     112735 TCTAGAATAAAAAATATTATCAGACATACTGTGAAAATTCCCTTGCTCTGTTAAACATTA 112794

Qy       1727 CATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAG 1786
              | || ||| | |||||||| |    || ||||   |  |  ||  ||  |||||  ||
Db     112795 TTTGGAAAATATTTTAAAAAGAAAAAAAAATCTTAGGGGGGCTAATAATTCTGACTTCAG 112854

Qy       1787 TGGACAGTG 1795
                |   |||
Db     112855 CTGTGTGTG 112863


RESULT 26
BX927412/c
LOCUS       BX927412              188520 bp    DNA     linear   HTG 13-MAY-2005
DEFINITION  Danio rerio clone CH211-143A8, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   BX927412
VERSION     BX927412.4  GI:64967770
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 188520)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 13, 2005 this sequence version replaced gi:45822745.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC143A8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 186450 bases at least Q40
            Consensus quality: 186777 bases at least Q30
            Consensus quality: 187184 bases at least Q20
            Insert size: 187920; sum-of-contigs
            Insert size: 181420; 27.5% error; agarose-fp
            Quality coverage: 7.23x in Q20 bases; sum-of-contigs Quality
            coverage: 7.60x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    23128: contig of 23128 bp in length
            *   23129    23228: gap of 100 bp
            *   23229    97391: contig of 74163 bp in length
            *   97392    97491: gap of 100 bp
            *   97492   107937: contig of 10446 bp in length
            *  107938   108037: gap of 100 bp
            *  108038   167507: contig of 59470 bp in length
            *  167508   167607: gap of 100 bp
            *  167608   169634: contig of 2027 bp in length
            *  169635   169734: gap of 100 bp
            *  169735   183842: contig of 14108 bp in length
            *  183843   183942: gap of 100 bp
            *  183943   188520: contig of 4578 bp in length.
FEATURES             Location/Qualifiers
     source          1. .188520
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-143A8"
                     /clone_lib="CHORI-211"
     misc_feature    1. .23128
                     /note="assembly_fragment:00412
                     fragment_chain:1"
     misc_feature    23229. .97391
                     /note="assembly_fragment:01532
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    97492. .107937
                     /note="assembly_fragment:00103
                     fragment_chain:1"
     misc_feature    108038. .167507
                     /note="assembly_fragment:00696
                     fragment_chain:1"
     misc_feature    167608. .169634
                     /note="assembly_fragment:00033
                     fragment_chain:1"
     misc_feature    169735. .183842
                     /note="assembly_fragment:00226
                     fragment_chain:1"
     misc_feature    183943. .188520
                     /note="assembly_fragment:00049
                     fragment_chain:1"
ORIGIN

  Query Match              2.6%;  Score 51.8;  DB 14;  Length 188520;
```

```
  Best Local Similarity   47.2%;  Pred. No. 0.55;
  Matches  158;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy       1434 GCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGT 1493
             | || |   |  ||  || ||||     |   || || ||     ||   |||||    |
Db     108439 GGTTATAGTGTTTATTTGCTTTAAGAAAATGCAGTCTATTTTTGGACCACTAAAAATATT 108380

Qy       1494 ATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTA 1553
             | ||    |||  |   |    |||||  |   ||  |||  |    ||||  | | | ||
Db     108379 AATGTTTTATGACAGTTAATCATCTTTAGTTACCAAATCTTGTTATCAAAGTAATAAATA 108320

Qy       1554 GATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATT 1613
               | |   | |  |   |   | | |   |   | |  | || |    || |
Db     108319 AAATAAAGTTACTGAAACAAAAAAGTTACTGTTTAATTTTTTAAGTAAAATGTTAACAAT 108260

Qy       1614 AGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATA 1673
             |  |||  || | ||  |   |   || | |||   |    ||  | ||||   | || |
Db     108259 AAATAATTAATATAATATTGTAAAAATATCATTTTCATTTTAATTTAGTTTCCATAAAAA 108200

Qy       1674 TAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAG 1733
              ||||| || |||||  |   || |  |  |  | |||| | |||  |    |  ||  |
Db     108199 AAAAAAAAAAAAAAAATATATATATATATATATATATTTTTTTTTTTTTTTTTTTG 108140

Qy       1734 AATCTATTAAAAAGTATTTGAATTCTTTATGATGA 1768
               |  |||  | || ||| |  ||| ||| | || |
Db     108139 TAAATATACATAAATATATTTATTTTTTGTAATAA 108105


RESULT 27
AC092392/c
LOCUS       AC092392               98286 bp    DNA     linear   PRI 15-DEC-2001
DEFINITION  Homo sapiens chromosome 1 clone RP4-544H6, complete sequence.
ACCESSION   AC092392 AL109625
VERSION     AC092392.2  GI:17861066
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 98286)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 98286)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 98286)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Dec 15, 2001 this sequence version replaced gi:14589610.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: SC
            -------------- Project Information
                Center project name: chr-1
                Center clone name: RP4-544H6 (sc0188)
            -------------- Summary Statistics
                Sequencing vector: plasmid; 22% of reads
                Sequencing vector: plasmid; L08752; 78% of reads
                Chemistry: Dye-terminator ET; 32% of reads
                Chemistry: Dye-terminator Big Dye; 68% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 98273 bases at least Q40
                Consensus quality: 98278 bases at least Q30
                Consensus quality: 98286 bases at least Q20
                Insert size: 98287; sum-of-contigs
                Quality coverage: 9.1x in Q20 bases; sum-of-contigs
            --------------.
            Overlapping Sequences:
                5':  RP5-1043L3 (UWGC:sc0545) AC099062
                3':  RP4-671I22 (UWGC:sc0205) AL356298
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
```

BglII		EcoRI		HindIII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
6023	5830	8696	8978	1849	1934
2067	2146	6101	6394	6382	6210
7084	7102	6972	6986	512	<800
2144	2146	6986	6986	449	<800
1462	1431	11036	10738	1792	1934
2877	2772	826	827	7699	7751
252	<800	1438	1380	3248	3382
547	<800	2212	2194	6225	6210
815	883	1590	1521	9	<800
308	<800	3225	3154	720	726
515	<800	3261	3312	1705	1735
1321	1320	671	<800	2231	2162
11034	10897	2171	2194	83	<800
4803	4847	3889	3911	298	<800

1643	1622	2401	2382	307	<800
2140	2146	742	827	3417	3382
2925	2947	8122	8089	2700	2801
4658	4573	988	974	262	<800
3569	3631	2263	2194	1407	1374
3936	4009	825	827	5580	5488
653	<800	7300	7287	2031	2162
18	<800	1900	1898	2574	2646
1053	1111	4808	4762	2085	2162
1135	1111	2669	2653	897	907
50	<800	2712	2746	202	<800
113	<800	6452	6394	706	<800
2617	2615	2160	2194	3060	3207
689	<800	3777	3706	3662	3618
402	<800	792	<800	1702	1655
7554	7725			250	<800
592	<800			1010	1005
5487	5529			333	<800
1643	1622			233	<800
6052	6239			1223	1200
896	883			4551	4630
5326	5322			1948	1934
1036	1030			4120	3964
6826	6811			7902	8074
1378	1320			5161	5086
599	<800			543	<800
2743	2615			6163	6210
				5260	5086
				1692	1655
				2232	2162
				451	<800
				119	<800

```
FEATURES             Location/Qualifiers
     source          1. .98286
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP4-544H6"
                     /clone_lib="RPCI human PAC library 4"
ORIGIN
```

```
  Query Match             2.6%;  Score 51.2;  DB 8;  Length 98286;
  Best Local Similarity   73.9%;  Pred. No. 0.81;
  Matches   65;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         55 TATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAA 114
              | |  ||||||||  | || ||| |||| |||||||| | |||||||  || || |  ||
Db       3689 TGTTTTTCAGGACTCTTGACTGCTAGTGTCAGAAACTAATTTCAAACTGACATATGCAAA 3630

Qy        115 AAGGAAAATATATTGGCTCATGTAACCT 142
              |  |||||| |||||| || |||||| |
Db       3629 AGAGAAAATGTATTGGTTCTTGTAACTT 3602
```

```
RESULT 28
AC099062
LOCUS       AC099062               101840 bp    DNA     linear   PRI 19-OCT-2002
DEFINITION  Homo sapiens chromosome 1 clone RP5-1043L3, complete sequence.
ACCESSION   AC099062 AL162737
VERSION     AC099062.2  GI:24137515
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 101840)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 101840)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 101840)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Oct 19, 2002 this sequence version replaced gi:16799020.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: SC
            -------------- Project Information
                Center project name: chr-1
                Center clone name: RP5-1043L3 (sc0545)
            ------------- Summary Statistics
                Sequencing vector: plasmid; 59% of reads
                Sequencing vector: plasmid; L08752; 41% of reads
                Chemistry: Dye-terminator ET; 67% of reads
                Chemistry: Dye-terminator Big Dye; 33% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 101791 bases at least Q40
                Consensus quality: 101836 bases at least Q30
                Consensus quality: 101840 bases at least Q20
                Insert size: 101840; sum-of-contigs
                Quality coverage: 8.5x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
                5': RP4-544H6 (UWGC:sc0188) AC092392, 25483-bp overlap
                3': RP5-1052I5 AL139139, 2000-bp overlap
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
```

All manually edited bases have been reduced to quality zero..
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

EcoRI		BglII		HindIII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
3363	3404	18776	19035	5271	5274
2184	2184	5671	5721	449	<800
8065	8113	4434	4469	512	<800
486	<800	11034	10925	2814	2830
106	<800	1321	1311	1247	1210
7605	7665	515	<800	4695	4715
6986	7005	308	<800	3373	3262
6972	7005	815	816	4101	4054
13043	12898	547	<800	307	<800
1585	1580	252	<800	298	<800
3216	3181	2877	2906	83	<800
1187	1187	1462	1441	2231	2220
220	<800	2144	2130	1705	1686
2897	2889	4681	4633	720	<800
3109	3181	2136	2130	9	<800
8409	8440	12670	12467	6225	6242
5818	5726	849	816	3248	3262
1702	1666	2983	3030	7699	7733
6859	7005	11716	11505	1764	1686
18143	18437	421	<800	1299	1294
13333	12898	5683	5721	855	872
		2562	2565	2287	2309
				1877	1866
				807	816
				2588	2649
				9960	9990
				329	<800
				2764	2799
				3477	3525
				620	<800
				1035	1018
				811	816
				5189	5100
				828	872
				5346	5274
				981	970
				879	872
				662	<800
				2519	2554
				3286	3368
				2695	2712
				7072	7095
				3584	3518
				457	<800
				5534	5568
				284	<800
				3513	3518
				3058	3085
				2204	2220
				862	872
				5062	5100
				4803	4715
				1502	1452
				5821	5758
				89	<800
				410	<800
				2305	2334

```
FEATURES             Location/Qualifiers
     source          1. .101840
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP5-1043L3"
                     /clone_lib="RPCI human PAC library 5"
     misc_feature    42195. .42333
                     /note="Single subclone region"
     misc_feature    43403. .43760
                     /note="Single subclone region"
     misc_feature    44068. .44144
                     /note="Single subclone region"
ORIGIN

  Query Match             2.6%;  Score 51.2;  DB 8;  Length 101840;
  Best Local Similarity   73.9%;  Pred. No. 0.8;
  Matches   65;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         55 TATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAA 114
              | |  ||||||||  | || ||| |||| |||||||| | |||||||  || || |  ||
```

```
Db      21795 TGTTTTTCAGGACTCTTGACTGCTAGTGTCAGAAACTAATTTCAAACTGACATATGCAAA 21854

Qy        115 AAGGAAAATATATTGGCTCATGTAACCT 142
              |  |||||| |||||| || |||||| |
Db      21855 AGAGAAAATGTATTGGTTCTTGTAACTT 21882


RESULT 29
AE014842/c
LOCUS       AE014842               253151 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 11 section 7 of 8 of the
            complete sequence.
ACCESSION   AE014842 AE014186
VERSION     AE014842.1  GI:23496321
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 253151)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 253151)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .253151
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="11"
     repeat_region   complement(73. .101)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   173. .279
                     /rpt_type=tandem
                     /rpt_unit="(taaaaa)n"
     repeat_region   complement(302. .328)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(393. .466)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   393. .442
                     /rpt_type=tandem
     repeat_region   440. .466
                     /rpt_type=tandem
     repeat_region   545. .578
                     /rpt_type=tandem
     repeat_region   complement(545. .578)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(580. .673)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   762. .795
                     /rpt_type=tandem
     repeat_region   complement(762. .795)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(869. .955)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(956. .975)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(983. .1040)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1094. .1121
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(1149. .1221)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1239. .1281
                     /rpt_type=tandem
     repeat_region   complement(1243. .1319)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(1323. .1379)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1571. .1605)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1621. .1699)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1820. .1872)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1921. .2010)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2105. .2135
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(2135. .2176)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(2226. .2248)
                     /rpt_type=tandem
                     /rpt_unit="(taaaa)n"
     repeat_region   complement(2289. .2362)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2372. .2396
                     /rpt_type=tandem
     repeat_region   complement(2372. .2396)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(2402. .2424)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(2438. .2620)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2636. .2671
                     /rpt_type=tandem
     repeat_region   complement(2636. .2671)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   2680. .2837
                     /rpt_type=tandem
                     /rpt_unit="(tataa)n"
     repeat_region   complement(2763. .2865)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(2932. .2954)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2970. .3008
```

```
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(3028. .3145)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(3189. .3238)
                     /rpt_type=tandem
                     /rpt_unit="(tataa)n"
     repeat_region   complement(3250. .3403)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   3425. .3520
                     /rpt_type=tandem
     repeat_region   complement(3425. .3520)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   3556. .3580
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(3584. .3641)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(3649. .3809)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(3842. .3884)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   3845. .3884
                     /rpt_type=tandem
     repeat_region   3897. .3944
                     /rpt_type=tandem
                     /rpt_unit="(tataa)n"
     repeat_region   complement(3946. .3999)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   4000. .4024
                     /rpt_type=tandem
     repeat_region   complement(4000. .4024)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(4026. .4104)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   4163. .4226
                     /rpt_type=tandem
                     /rpt_unit="(taaaa)n"
     repeat_region   4232. .4267
                     /rpt_type=tandem
     repeat_region   complement(4232. .4267)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(4339. .4359)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   4380. .4407
                     /rpt_type=tandem
     repeat_region   complement(4380. .4407)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(4408. .4575)
                     /note="AT_rich"
                     /rpt_type=tandem

  Query Match             2.6%;  Score 51.2;  DB 2;  Length 253151;
  Best Local Similarity   45.7%;  Pred. No. 0.72;
  Matches  215;  Conservative    0;  Mismatches  253;  Indels    2;  Gaps    1;

Qy       1313 TTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTA 1372
             | |  |||||  |||  | || | | |    |||| |   ||  |  |||  |   |
Db     178545 TAAAGGCATATATATAATAAAATAAATATATATTTTTTGTATCGTTTTATATATTTATTT 178486

Qy       1373 CACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTG 1432
```

```
             | |   |   | | |   |   |  |||  |  || |||  ||||||   |  |
Db     178485 TATATTTTTTTTTTTTTTTTAATTGTATATTTTTTTTTATATAATTATTATATAATATA 178426

Qy       1433 TGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGG 1492
             |  || |     | | || | ||||||  |   |   | |    | | |||||  |
Db     178425 TAATTATGGGAAGAACATATATTATATATATACATTATTTATATTAATATATATATATAT 178366

Qy       1493 TATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATT 1552
              ||     ||    ||| |    | |||  |  | |  |  |  || |     | |  ||
Db     178365 AATATTATAAATAAATATATTTTTATTTATTTTATATATTCCTCATTTTTTATAATTTTT 178306

Qy       1553 AGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCAT 1612
              |  |     ||||||   |  | | |  || | | |      || | ||   |||  |||
Db     178305 ATTTCTTTAGAAATATCTTTTTATTTATATATATATTTATTATTTTTATTTTGAAAACAT 178246

Qy       1613 TAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAAT 1672
              |    |  || || |  || ||| |       ||  ||||  ||  || |||   |  ||
Db     178245 CATAATAATAATAATATTTAAAATATTAAACCTTTTAAGACGAA--TGTTTTGTATTTAT 178188

Qy       1673 ATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGA 1732
              |||  | | | |||  |    |||  | |||  |  |  |||  |   |   || |
Db     178187 TATTTTTTAATATATATAAAAAAATATAAAATAAAATAAGTATTTTTTTTTAGGTTTTA 178128

Qy       1733 GAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACT 1782
              |||    || |  |||||   || |||||    | ||   ||  || |
Db     178127 TTATCATCAAATATATATTTTTTTTTTTTATATGAAAAAAATTCACGAGT 178078
```

```
RESULT 30
AC111006
LOCUS       AC111006              123019 bp    DNA     linear   PRI 29-MAY-2002
DEFINITION  Homo sapiens BAC clone RP11-761E21 from 4, complete sequence.
ACCESSION   AC111006
VERSION     AC111006.5  GI:21070789
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 123019)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 123019)
  AUTHORS   Isak,A., Meyer,R. and Creason,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-761E21
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 123019)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 123019)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 123019)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 123019)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 22, 2002 this sequence version replaced gi:20146839.
```

```
            ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            ---------- Summary Statistics
                Center project name: H_NH0761E21
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-58N10; the clone sequenced
            to the right is RP11-63A11, 2000 bp overlap.  Actual start of this
            clone is at base position 1 of RP11-761E21; actual end is at base
            position 77234 of RP11-63A11.

            The region from 55547 to 55670 is covered only by a pcr product
            from clone DNA. Polymorphisms exist between AC113151 and AC111006.
            Data from AC113151 and AC007799 was used to finish AC111006.
FEATURES             Location/Qualifiers
     source          1. .123019
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-761E21"
                     /clone_lib="RPCI-11"
     repeat_region   19. .392
                     /rpt_family="L2"
     repeat_region   393. .693
                     /rpt_family="Alu"
     repeat_region   694. .923
                     /rpt_family="L2"
     repeat_region   993. .1009
                     /rpt_family="L1"
     repeat_region   1010. .1140
                     /rpt_family="Alu"
     repeat_region   1141. .1203
                     /rpt_family="L1"
     repeat_region   1295. .1600
                     /rpt_family="Alu"
     repeat_region   1769. .1896
                     /rpt_family="L1"
     repeat_region   1923. .2198
                     /rpt_family="Alu"
     repeat_region   2205. .2239
                     /rpt_family="(GAAAA)n"
     repeat_region   2650. .2701
                     /rpt_family="L2"
     repeat_region   3704. .4984
                     /rpt_family="L1"
     repeat_region   4985. .5448
                     /rpt_family="MER2_type"
     repeat_region   5449. .5524
                     /rpt_family="MER2_type"
     repeat_region   5512. .5604
                     /rpt_family="MER2_type"
     repeat_region   5605. .6414
                     /rpt_family="L1"
     repeat_region   6415. .6521
                     /rpt_family="MER2_type"
     repeat_region   6524. .6831
                     /rpt_family="Alu"
     repeat_region   6873. .7093
                     /rpt_family="L1"
     repeat_region   7165. .7531
                     /rpt_family="Mariner"
     repeat_region   7749. .7861
                     /rpt_family="MER1_type"
     repeat_region   8045. .8197
                     /rpt_family="L1"
     repeat_region   8198. .8624
                     /rpt_family="MaLR"
     repeat_region   8625. .8824
                     /rpt_family="L1"
     repeat_region   9046. .9327
                     /rpt_family="Alu"
     repeat_region   9510. .9530
                     /rpt_family="AT_rich"
     repeat_region   9731. .9898
                     /rpt_family="MER2_type"
     repeat_region   9899. .10032
                     /rpt_family="MER1_type"
     repeat_region   10033. .10327
                     /rpt_family="Alu"
     repeat_region   10328. .10408
                     /rpt_family="MER1_type"
     repeat_region   10409. .10582
                     /rpt_family="MER2_type"
     repeat_region   10595. .10908
                     /rpt_family="Alu"
     repeat_region   11430. .11547
                     /rpt_family="Alu"
     repeat_region   11664. .11753
                     /rpt_family="L2"
     repeat_region   11850. .13322
                     /rpt_family="L1"
     repeat_region   13323. .13635
                     /rpt_family="Alu"
     repeat_region   13636. .13666
                     /rpt_family="L1"
     repeat_region   13826. .14116
                     /rpt_family="Alu"
     repeat_region   16039. .16162
                     /rpt_family="MIR"
     repeat_region   22848. .22955
                     /rpt_family="MER1_type"
     repeat_region   25267. .25393
                     /rpt_family="MIR"
     repeat_region   26158. .26283
                     /rpt_family="L2"
     repeat_region   26305. .26353
                     /rpt_family="L2"
     repeat_region   27372. .27402
                     /rpt_family="L2"
     repeat_region   27515. .27537
                     /rpt_family="(A)n"
```

```
     repeat_region   27897. .28177
                     /rpt_family="Alu"
     repeat_region   29544. .29564
                     /rpt_family="AT_rich"
     repeat_region   29566. .29860
                     /rpt_family="Alu"
     repeat_region   30550. .30673
                     /rpt_family="Alu"
     repeat_region   30677. .30714
                     /rpt_family="(CAAAA)n"
     repeat_region   31380. .31475

  Query Match             2.5%;  Score 50.6;  DB 8;  Length 123019;
  Best Local Similarity   65.5%;  Pred. No. 1.1;
  Matches   74;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         30 AGAGTGAAAAATATATTAACGTGCATATAGTTCAGGACCATGGATTGCAAGTGACAGAAA 89
              ||   || |  || | |       | | ||  |   |||  | | ||||||| ||||||||
Db      68191 AGGAAGAGATTTAAAGTTTTAGTCTTTTACATTCTGACTCTTGGTTGCAAGAGACAGAAA 68250

Qy         90 CTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATATTGGCTCATGTAACCT 142
                ||| ||||||||   |||||  ||||| |||| ||||||||||  ||| |
Db      68251 TCCAACTCAAACCAGATTAAGTACAAAGGGAAATTTATTGGCTCATTCAACTT 68303


RESULT 31
AC108123/c
LOCUS       AC108123              191481 bp    DNA     linear   HTG 25-JAN-2002
DEFINITION  Homo sapiens chromosome 16 clone RP11-565I1, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC108123
VERSION     AC108123.1  GI:18369969
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 191481)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191481)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 623518
            Center clone name: RPCI-11_565I1
            -------
            Summary Statistics
            Consensus quality: 182477 bases at least Q40
            Consensus quality: 187876 bases at least Q30
            Consensus quality: 188750 bases at least Q20
            Estimated insert size: 192000; agarose-fp estimation
            Estimated insert size: 190581; sum-of-contigs estimation
            Quality coverage:  6.71 in Q20 bases; agarose-fp estimation
            Quality coverage:  6.76 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1230: contig of 1230 bp in length
            *     1231     1330: gap of unknown length
            *     1331     3606: contig of 2276 bp in length
            *     3607     3706: gap of unknown length
            *     3707     6459: contig of 2753 bp in length
            *     6460     6559: gap of unknown length
            *     6560    10289: contig of 3730 bp in length
            *    10290    10389: gap of unknown length
            *    10390    16864: contig of 6475 bp in length
            *    16865    16964: gap of unknown length
            *    16965    27807: contig of 10843 bp in length
            *    27808    27907: gap of unknown length
            *    27908    39098: contig of 11191 bp in length
            *    39099    39198: gap of unknown length
            *    39199    76534: contig of 37336 bp in length
            *    76535    76634: gap of unknown length
            *    76635   121347: contig of 44713 bp in length
            *   121348   121447: gap of unknown length
            *   121448   191481: contig of 70034 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191481
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-565I1"
                     /clone_lib="RPCI human BAC library 11"
     gap             1231. .1330
                     /estimated_length=unknown
     gap             3607. .3706
                     /estimated_length=unknown
     gap             6460. .6559
                     /estimated_length=unknown
     gap             10290. .10389
                     /estimated_length=unknown
     gap             16865. .16964
                     /estimated_length=unknown
     gap             27808. .27907
                     /estimated_length=unknown
     gap             39099. .39198
                     /estimated_length=unknown
     gap             76535. .76634
                     /estimated_length=unknown
     gap             121348. .121447
                     /estimated_length=unknown
ORIGIN

  Query Match             2.5%;  Score 49.8;  DB 14;  Length 191481;
  Best Local Similarity   53.3%;  Pred. No. 1.5;
  Matches  105;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy       1552 TAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCA 1611
              || | |   ||| ||| ||  || | |  |||| | |    || ||| |||   || | |
Db      10625 TATAATATATGATATATACAAATATATATATTTATATAAATACATATTATATATAAATAA 10566

Qy       1612 TTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAA 1671
              ||  |||  |   || |    ||| ||| ||     ||  | ||     ||| | | ||
Db      10565 ATATATAAATATATAATATATAAATATATTTATAATTAATTATAAATATATTTAGTTAAA 10506

Qy       1672 TATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTG 1731
              |||||| |||  || |   ||  ||    |  |  | |||  | | |||| | |    |
Db      10505 TATAAATATATTAATATATAATTATAATTATATTTATATAAATTTATAATTAAGATTATA 10446

Qy       1732 AGAATCTATTAAAAAGT 1748
              |  || ||||| | | |
Db      10445 ATTATATATTATATAAT 10429


RESULT 32
BD038954
LOCUS       BD038954                431 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Sequence tag and encoded human protein.
ACCESSION   BD038954
VERSION     BD038954.1  GI:22580696
```

```
KEYWORDS    JP 2001269182-A/15200.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
  TITLE     Sequence tag and encoded human protein
  JOURNAL   Patent: JP 2001269182-A 15200 02-OCT-2001;
            GENSET
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001269182-A/15200
            PD   02-OCT-2001
            PF   24-FEB-2000 JP 2000118773
            PR   26-FEB-1999 US    60/122487
            PI   JEAN BAPUTIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
            PI   JORDAN
            PC   C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
             C12N5/10,
            PC   C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
              G06F15/40
            CC
            FH   Key             Location/Qualifiers.
FEATURES              Location/Qualifiers
     source           1. .431
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             2.5%;  Score 49.4;  DB 6;  Length 431;
  Best Local Similarity   71.4%;  Pred. No. 3.9;
  Matches   65;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         70 TGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATATTG 129
              | | ||| |||||||||||||| ||| |||||  | | || | ||||| || ||||||| |
Db        231 TTGGTTGTAAGTGACAGAAACCCAACTCAAATTAGCTTAGGCCAAAAAGAGAATATATGG 290

Qy        130 GCTCATGTAACCTTCTCACAGAGAGGGCAGG 160
              | ||||| ||||    |  ||| | || | |
Db        291 GTTCATGGAACCCAAACTAAGAAAAGGGACG 321


RESULT 33
AX903421
LOCUS       AX903421                431 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 19284 from Patent EP1033401.
ACCESSION   AX903421
VERSION     AX903421.1  GI:40058378
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
  TITLE     Expressed sequence tags and encoded human proteins
  JOURNAL   Patent: EP 1033401-A 19284 06-SEP-2000;
            Genset (FR)
FEATURES              Location/Qualifiers
     source           1. .431
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             2.5%;  Score 49.4;  DB 6;  Length 431;
  Best Local Similarity   71.4%;  Pred. No. 3.9;
  Matches   65;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         70 TGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATATTG 129
              | | ||| |||||||||||||| ||| |||||  | | || | ||||| || ||||||| |
Db        231 TTGGTTGTAAGTGACAGAAACCCAACTCAAATTAGCTTAGGCCAAAAAGAGAATATATGG 290

Qy        130 GCTCATGTAACCTTCTCACAGAGAGGGCAGG 160
              | ||||| ||||    |  ||| | || | |
Db        291 GTTCATGGAACCCAAACTAAGAAAAGGGACG 321


RESULT 34
CS083840/c
LOCUS       CS083840                1428 bp    DNA     linear   PAT 18-MAY-2005
DEFINITION  Sequence 23 from Patent WO2005040377.
ACCESSION   CS083840
VERSION     CS083840.1  GI:66349967
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mermod,N., Girod,P.A., Bucher,P., Nguyen,D.Q., Calabrese,D.,
            Saugy,D. and Puttini,S.
  TITLE     High efficiency gene transfer and expression in mammalian cells by
            a multiple transfection procedure of mar sequences
  JOURNAL   Patent: WO 2005040377-A 23 06-MAY-2005;
            Selexis S.A. (CH)
FEATURES              Location/Qualifiers
     source           1. .1428
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
                      /note="MAR of human chromosome 2, genomic contig; 17840365
                      to 17841792"
ORIGIN

  Query Match             2.5%;  Score 49.4;  DB 6;  Length 1428;
  Best Local Similarity   46.0%;  Pred. No. 3.4;
  Matches  210;  Conservative    0;  Mismatches  241;  Indels    6;  Gaps    1;

Qy       1310 ATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCAT 1369
              | | || |  ||| | ||  | |||  | | |||| ||  |  | | |   |||  | |
Db       1395 AATATAATAAATATTAATAATTAATAAATTAATAAATTTAATTAATATATTATATATAAA 1336

Qy       1370 CTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAG 1429
              || | |  |||   | | ||  |  |  |   | |  ||||| ||      ||  | |
Db       1335 ATATATATTATAATATATTATATATAAAATATATATTATAATATATTATATATAAAATAT 1276

Qy       1430 CTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAG 1489
              | |  |  |  |  || ||   |||||   |||     |   ||| ||||||||
Db       1275 ATATTATAATATATTATATATAAAATATATATTATAA------TATATTATATATAAAAT 1222

Qy       1490 GGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTC 1549
              ||||     || | ||| | ||    | | |    |  | | |||| ||||  ||
Db       1221 ATATATTATAATATATTATATATAAAATATATATTATAATATATTATATATAAAATATAT 1162

Qy       1550 ATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGT 1609
              |||| | |   | | ||| |           || |      | |  |||       ||
Db       1161 ATTATAATATATTATATATAAAATATATATTATATTATATATAAAATATATATTATAATA 1102

Qy       1610 CATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTC 1669
              ||||  ||   || | | | || ||| |||   ||   || | |   |   | |   |
Db       1101 TATTATATATAAAATATATATTATAATATATTATATATAAAATATATATTATAATATATT 1042

Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
              |     ||||||   | | ||   || |   |     | ||  |   | ||  | | |
Db       1041 ATATATAAAATATATATTATAATATATATTATAATATATTATATATAAAATATATATT 982

Qy       1730 TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGAT 1766
              |  || |||   | | ||| | |  | |  |  ||
Db        981 ATAATATATATATTATAATATATTATATATAAAATAT 945
```

```
RESULT 35
AC104069/c
LOCUS       AC104069              155106 bp    DNA     linear   PRI 16-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-279B11 from 2, complete sequence.
ACCESSION   AC104069 AC027210
VERSION     AC104069.3  GI:18855142
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 155106)
  AUTHORS   Scott,K., Kozlowicz,A. and Creason,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-279B11
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 155106)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 155106)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 155106)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 21, 2002 this sequence version replaced gi:17921249.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0279B11
                Drafting Center: WIBR
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-362A12; the clone sequenced
            to the right is RP11-507C12.  Actual start of this clone is at base
            position 1 of RP11-279B11; actual end is at base position 155106 of
            RP11-279B11.

            Data from AC104409 and AC093040 was used to finish this clone,
            AC104069. Polymorphisms have been identified between AC104409 and
            AC104069.

            The sequence of AC027210 has been incorporated into AC104069.
FEATURES             Location/Qualifiers
     source          1. .155106
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-279B11"
                     /clone_lib="RPCI-11"
ORIGIN

  Query Match             2.5%;  Score 49.4;  DB 8;  Length 155106;
  Best Local Similarity   46.0%;  Pred. No. 1.9;
  Matches  210;  Conservative    0;  Mismatches  241;  Indels    6;  Gaps    1;

Qy       1310 ATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCAT 1369
              | | || |  ||| | ||  | |||  | | |||| ||  |  | | |   |||  | |
Db      99925 AATATAATAAATATTAATAATTAATAAATTAATAAATTTAATTAATATATTATATATAAA 99866

Qy       1370 CTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAG 1429
              || | |  |||   | | ||  |  |  |  | |  ||||| ||      ||  | |
Db      99865 ATATATATTATAATATATTATATATAAAATATATATTATAATATATTATATATAAAATAT 99806

Qy       1430 CTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAG 1489
              | |  |  |  |   || ||    |||||    |||     |   ||| ||||||||
Db      99805 ATATTATAATATATTATATATAAAATATATATTATAA------TATATTATATATAAAAT 99752

Qy       1490 GGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTC 1549
              ||||     || | ||| | ||    | | |    |  | | |||| ||||  ||
Db      99751 ATATATTATAATATATTATATATAAAATATATATTATAATATATTATATATAAAATATAT 99692

Qy       1550 ATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGT 1609
              |||| | |   | | ||| |           || |      | |  |||       ||
Db      99691 ATTATAATATATTATATATAAAATATATATTATATTATATATAAAATATATATTATAATA 99632

Qy       1610 CATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTC 1669
              ||||  ||   || | | | || ||| |||   ||   ||  | |   |    |  |
Db      99631 TATTATATATAAAATATATATTATAATATATTATATATAAAATATATATTATAATATATT 99572

Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
              |     ||||||   |  | ||  || |   |    | ||  |   | ||  |  | |
Db      99571 ATATATAAAATATATATTATAATATATATTATTATAATATATTATATATAAAATATATATT 99512

Qy       1730 TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGAT 1766
              |  || |||   | | ||| | |  | |  |  ||
Db      99511 ATAATATATATATTATAATATATTATATATAAAATAT 99475


RESULT 36
AL450263
LOCUS       AL450263              157671 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP13-225O21 on chromosome 9 Contains
            the 3' end of the RALGPS1A gene for Ral guanine nucleotide exchange
            factor RalGPS1A (RALGEF2, KIAA0351), two novel genes, the gene for
            a novel protein (DKFZp761J1523), a density-regulated protein
            (DENR)(DRP, DRP1, SMAP-3) pseudogene and two CpG islands, complete
            sequence.
ACCESSION   AL450263
VERSION     AL450263.15  GI:13234989
KEYWORDS    HTG; CpG island; DENR; DFKZp761J1523; DRP; DRP1; KIAA0351; RALGEF2;
            RALGPS1A; SMAP-3.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 157671)
  AUTHORS   Babbage,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 5, 2001 this sequence version replaced gi:13184383.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP13-225O21 is from the library RPCI-13.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .157671
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP13-225O21"
                     /clone_lib="RPCI-13.1"
     misc_feature    1
                     /note="Clone_left_end: RP13-225O21"
     misc_feature    1749. .1858
                     /note="Sequence from overlapping clones bA342E24
                     (AL357623) and bA131B16 (AL354705). Assembly confirmed by
                     restriction digest."
     gene            join(AL160169.12:31066. .31267,AL160169.12:78517. .78638,
                     AL160169.12:82136. .82243,AL160169.12:93987. .94037,
                     AL356862.10:28698. .28781,AL356862.10:44310. .44399,
                     AL356862.10:47114. .47206,AL356862.10:63497. .63623,
                     AL357623.14:16855. .16992,AL357623.14:19465. .19558,
                     AL357623.14:25501. .25568,8456. .8582,9839. .9996,
                     12380. .12431,25496. .25601,25992. .26084,26309. .26414,
                     28116. .28207,32105. .36529)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-001"
     mRNA            join(AL160169.12:31066. .31267,AL160169.12:78517. .78638,
                     AL160169.12:82136. .82243,AL160169.12:93987. .94037,
                     AL356862.10:28698. .28781,AL356862.10:44310. .44399,
                     AL356862.10:47114. .47206,AL356862.10:63497. .63623,
                     AL357623.14:16855. .16992,AL357623.14:19465. .19558,
                     AL357623.14:25501. .25568,8456. .8582,9839. .9996,
                     12380. .12431,25496. .25601,25992. .26084,26309. .26414,
                     28116. .28207,32105. .36529)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-001"
                     /note="match: ESTs: AA121710.1 AA126485.1 AA126611.1
                     AA284968.1 AA293632.1 AA354061.1 AA446069.1 AA459215.1
                     AA478123.1 AA478986.1 AA815347.1 AI183582.1 AI582306.1
                     AI613219.1 AI656043.1 AI697716.1 AI909150.1 AI909967.1
                     AL600005.1 AU119994.1 AU134174.1 AV753402.1 AW118605.1
                     AW137800.1 AW138212.1 AW750651.1 AW842484.1 AW959164.1
                     AW968411.1 AW976742.1 BF196253.1 BF951871.1 BG433408.1
                     BG984586.1 BI050462.1 BI598980.1 BI916008.1 BM509596.1
                     BM681237.1 BM685093.1 BM726588.1 BM741025.1 BM745776.1
                     BM746243.1 BM753571.1 BM764434.1 BM765703.1 BM765713.1
                     BM766542.1 BM772138.1 BM772427.1 BM932077.1 BM983757.1
                     BQ023410.1 BQ185492.1 BQ187845.1 BQ188851.1 BQ228362.1
                     BQ324015.1 CA392005.1 CA412508.1 F07861.1 H12126.1
                     H18534.1 H19388.1 H55229.1 W28684.1 W28685.1 W28701.1
                     match: cDNAs: AB002349.1"
     CDS             join(AL160169.12:78582. .78638,AL160169.12:82136. .82243,
                     AL160169.12:93987. .94037,AL356862.10:28698. .28781,
                     AL356862.10:44310. .44399,AL356862.10:47114. .47206,
                     AL356862.10:63497. .63623,AL357623.14:16855. .16992,
                     AL357623.14:19465. .19558,AL357623.14:25501. .25568,
                     8456. .8582,9839. .9996,12380. .12431,25496. .25601,
                     25992. .26084,26309. .26414,28116. .28207,32105. .32134)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-001"
                     /standard_name="OTTHUMP00000022161"
                     /note="match: proteins: O15059"
                     /codon_start=1
                     /protein_id="CAI41461.1"
                     /db_xref="GI:57209377"
                     /db_xref="GOA:Q5VXM4"
                     /db_xref="InterPro:IPR001849"
                     /db_xref="InterPro:IPR001895"
                     /db_xref="UniProt/TrEMBL:Q5VXM4"
                     /translation="MYKRNGLMASVLVTSATPQGSSSSDSLEGQSCDYASKSYDAVVF
                     DVLKVTPEEFASQITLMDIPVFKAIQPEELASCGWSKKEKHSLAPNVVAFTRRFNQVS
                     FWVVREILTAQTLKIRAEILSHFVKIAKKLLELNNLHSLMSVVSALQSAPIFRLTKTW
                     ALLNRKDKTTFEKLDYLMSKEDNYKRTREYIRSLKMVPSIPYLGIYLLDLIYIDSAYP
                     ASGSIMENEQRSNQMNNILRIIADLQVSCSYDHLTTLPHVQKYLKSVRYIEELQKFVE
                     DDNYKLSLRIEPGSSSPRLVSSKEDLAGPSAGSGSARFSRRPTCPDTSVAGSLPTPPV
                     PRHRKSHSLGNNMMCQLSVVESKSATFPSEKARHLLDDSVLESRSPRRGLALTSSSAV
                     TNGLSLGSSESSEFSEEMSSGLESPTGPCICSLGNSAAVPTMEGPLRRKTLLKEGRKP
                     ALSSWTRYWVILSGSTLLYYGAKSLRGTDRKHYKSTPGKKVSIVGWMVQLPDDPEHPD
                     IFQLNNPDKGNVYKFQTGSRFHAILWHKHLDDACKSNRPQVPANLMSFE"
     gene            join(AL160169.12:78517. .78638,AL160169.12:82136. .82243,
                     AL160169.12:93987. .94037,AL356862.10:28698. .28781,
                     AL356862.10:44310. .44399,AL356862.10:47114. .47206,
                     AL356862.10:63497. .63623,AL357623.14:16855. .16992,
                     AL357623.14:19465. .19558,AL357623.14:25501. .25568,
                     8456. .8582,24060. .24143,25496. .25601,25992. .26084,
                     26333. .26414,28116. .28207,30271. .31177)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-002"
     mRNA            join(AL160169.12:78517. .78638,AL160169.12:82136. .82243,
                     AL160169.12:93987. .94037,AL356862.10:28698. .28781,
                     AL356862.10:44310. .44399,AL356862.10:47114. .47206,
                     AL356862.10:63497. .63623,AL357623.14:16855. .16992,
                     AL357623.14:19465. .19558,AL357623.14:25501. .25568,
                     8456. .8582,24060. .24143,25496. .25601,25992. .26084,
                     26333. .26414,28116. .28207,30271. .31177)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-002"
                     /note="match: ESTs: AA121487.1 AA478789.1 AI627675.1
                     AW245646.1 AW250005.1 BE263432.1 BF510295.1 CA392004.1
                     match: cDNAs: AF221098.1"
     CDS             join(AL160169.12:78582. .78638,AL160169.12:82136. .82243,
                     AL160169.12:93987. .94037,AL356862.10:28698. .28781,
                     AL356862.10:44310. .44399,AL356862.10:47114. .47206,
                     AL356862.10:63497. .63623,AL357623.14:16855. .16992,
                     AL357623.14:19465. .19558,AL357623.14:25501. .25568,
                     8456. .8582,24060. .24143,25496. .25601,25992. .26084,
                     26333. .26414,28116. .28207,30271. .30366)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-002"
                     /standard_name="OTTHUMP00000022162"
                     /note="match: proteins: Q9NZ16"
                     /codon_start=1
                     /protein_id="CAI41460.1"
                     /db_xref="GI:57209376"
                     /db_xref="GOA:Q9NZ16"
```

```
                     /db_xref="InterPro:IPR001849"
                     /db_xref="InterPro:IPR001895"
                     /db_xref="UniProt/TrEMBL:Q9NZ16"
                     /translation="MYKRNGLMASVLVTSATPQGSSSSDSLEGQSCDYASKSYDAVVF
                     DVLKVTPEEFASQITLMDIPVFKAIQPEELASCGWSKKEKHSLAPNVVAFTRRFNQVS
                     FWVVREILTAQTLKIRAEILSHFVKIAKKLLELNNLHSLMSVVSALQSAPIFRLTKTW
                     ALLNRKDKTTFEKLDYLMSKEDNYKRTREYIRSLKMVPSIPYLGIYLLDLIYIDSAYP
                     ASGSIMENEQRSNQMNNILRIIADLQVSCSYDHLTTLPHVQKYLKSVRYIEELQKFVE
                     DDNYKLSLRIEPGSSSPRLVSSKEDLAGPSAGSGSARFSRRPTCPDTSVAGSLPTPPV
                     PRHRKSHSLGNNRGRLYATLGPNWRVPVRNSPRTRSCVYSPTGPCICSLGNSAAVPTM
                     EGPLRRKTLLKEGRKPALSSWTRYWVILSGSTLLYYGAKSLRGTDRKHVSIVGWMVQL
                     PDDPEHPDIFQLNNPDKGNVYKFQTGSRFHAILWHKHLDDACKSNRPQEAGAAPGPTG
                     TDSHEVDHLEGGAGKEAGPCA"
     misc_feature    17109. .17357
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-002"
                     /note="Sequence from overlapping clones bA342E24
                     (AL357623) and bA131B16 (AL354705). Assembly confirmed by
                     restriction digest."
     gene            join(23868. .23890,24060. .24143,25496. .25601,25992. .26084,
                     26309. .26332,26385. .26397)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-003"
     mRNA            join(23868. .23890,24060. .24143,25496. .25601,25992. .26084,
                     26309. .26332,26385. .26397)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-003"
                     /note="match: ESTs: AV653914.1"
     gene            join(<24110. .24143,25496. .25601,25992. .26084,26309. .26414,
                     28116. .28207,32105. .36531)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-005"
     mRNA            join(<24110. .24143,25496. .25601,25992. .26084,26309. .26414,
                     28116. .28207,32105. .36531)
                     /gene="RP13-225O21.1"
                     /locus_tag="RP13-225O21.1-005"
                     /note="match: cDNAs: BC037725.1"
     CDS             join(<24110. .24143,25496. .25601,25992. .26084,26309. .26414,

  Query Match             2.5%;  Score 49.4;  DB 8;  Length 157671;
  Best Local Similarity   71.4%;  Pred. No. 1.9;
  Matches   65;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         70 TGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATATTG 129
              | | ||| |||||||||||||| ||| |||||  | | || | ||||| || ||||||| |
Db      60652 TTGGTTGTAAGTGACAGAAACCCAACTCAAATTAGCTTAGGCCAAAAAGAGAATATATGG 60711

Qy        130 GCTCATGTAACCTTCTCACAGAGAGGGCAGG 160
              | ||||| ||||    |  ||| | || | |
Db      60712 GTTCATGGAACCCAAACTAAGAAAAGGGACG 60742


RESULT 37
AL354705/c
LOCUS       AL354705              161027 bp    DNA     linear   HTG 16-OCT-2000
DEFINITION  Homo sapiens chromosome 9 clone RP11-131B16, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AL354705
VERSION     AL354705.10  GI:10862755
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 17, 2000 this sequence version replaced gi:10834465.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA131B16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 160247 bases at least Q40
            Consensus quality: 160580 bases at least Q30
            Consensus quality: 160708 bases at least Q20
            Insert size: 160827; sum-of-contigs
            Insert size: 165550; agarose-fp
            Quality coverage: 6.63x in Q20 bases; sum-of-contigs Quality
            coverage: 6.69x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    67522: contig of 67522 bp in length
            *    67523    67622: gap of 100 bp
            *    67623   104112: contig of 36490 bp in length
            *   104113   104212: gap of 100 bp
            *   104213   161027: contig of 56815 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161027
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-131B16"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .67522
                     /note="assembly_fragment:00751
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    67623. .104112
                     /note="assembly_fragment:01022
                     fragment_chain:1"
     misc_feature    104213. .161027
                     /note="assembly_fragment:02056
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             2.5%;  Score 49.4;  DB 14;  Length 161027;
  Best Local Similarity   71.4%;  Pred. No. 1.9;
  Matches   65;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         70 TGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATATTG 129
              | | ||| |||||||||||||| ||| |||||  | | || | ||||| || ||||||| |
Db      81740 TTGGTTGTAAGTGACAGAAACCCAACTCAAATTAGCTTAGGCCAAAAAGAGAATATATGG 81681

Qy        130 GCTCATGTAACCTTCTCACAGAGAGGGCAGG 160
              | ||||| ||||    |  ||| | || | |
Db      81680 GTTCATGGAACCCAAACTAAGAAAAGGGACG 81650


RESULT 38
PFMAL4P3/c
LOCUS       PFMAL4P3              349751 bp    DNA     linear   INV 16-APR-2005
DEFINITION  Plasmodium falciparum MAL4P3.
ACCESSION   AL035476 AL844503
VERSION     AL035476.7  GI:23498195
```

```
KEYWORDS        .
SOURCE          Plasmodium falciparum 3D7
  ORGANISM      Plasmodium falciparum 3D7
                Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE       1
  AUTHORS       Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
                Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
                Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
                Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
                Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
                Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
                Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
                Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
                Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
                Lawson,D., Lennard,N., Line,A., Maddison,M., McLean,J., Mooney,P.,
                Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
                Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
                Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
                Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
                Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
                Barrell,B.G.
  TITLE         Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
  JOURNAL       Nature 419 (6906), 527-531 (2002)
   PUBMED       12368867
REFERENCE       2  (bases 1 to 349751)
  AUTHORS       Hamlin,N., Pain,A., Berriman,B., Hall,N., Bowman,S., Churcher,C.,
                Harris,B., Harris,D., Lawson,D., Quail,M. and Barrell,B.
  TITLE         Direct Submission
  JOURNAL       Submitted (24-FEB-1999) P.falciparum Genome Sequencing Consortium,
                The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
                CB10 1SA, UK
COMMENT         On Oct 3, 2002 this sequence version replaced gi:5763808.
                For more information about this sequence or the Malaria Project,
                see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .349751
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="4"
                     /clone="MAL4P3"
     gene            complement(join(2130. .3551,4472. .9988))
                     /gene="PFD0630c"
                     /note="synonym: VAR"
     CDS             complement(join(2130. .3551,4472. .9988))
                     /gene="PFD0630c"
                     /note="Similar to Plasmodium falciparum var, mal4p2.58
                     mal4p2.58, vaR SWALL:Q9U0G5 (EMBL:AL035475) (2277 aa)
                     fasta scores: E(): 1.5e-143, 52.77% id in 2486 aa, and to
                     Plasmodium falciparum variant-specific surface protein
                     var-7 SWALL:Q26034 (EMBL:L42636) (2182 aa) fasta scores:
                     E(): 1.9e-126, 46.72% id in 2412 aa"
                     /codon_start=1
                     /product="erythrocyte membrane protein 1 (PfEMP1)"
                     /protein_id="CAD49166.1"
                     /db_xref="GI:23498196"
                     /db_xref="GOA:Q8I1V7"
                     /db_xref="InterPro:IPR004258"
                     /db_xref="UniProt/TrEMBL:Q8I1V7"
                     /translation="MGPPSTAPDYSSAKHALDSIGKRVHAQVQNEAKQRSNGDLKGLL
                     TSATLSGGESAFTENPCELIKDKGDELLGDSGERHPCGNGSASEKRFSEVSGGECDDK
                     KIEGNGRNNGGACAPYRRLFLCNKNMEKMGRTSTTKHDLLLDVCMAANYEAQSLIRYH
                     DQYDATYPGSDFSMCTMLARSFADIGDIIRGKDLYLGKKKKKKTETERDQLESKLKKI
                     FGDIYNELTNGRNGVKDHYQDDNGGNYFQLREDWWTANRATVWKAITCKADTGNAYFR
                     PTCSNRQGPSQAHHYCRCNGDKPDDDKPNTDPPTYFDYVPQYLRWFEEWAEDFCRKKK
                     IYVGIVKTYCREKYKSGNEPRYCSRNGYDCTKTKRAIGKYRMGNQCISCLYACNPYVD
                     WINNQKEQFDKQKKKYTDEINEASRSSRRQKRGARSTGSSSNYDGYESKFYNILKDDY
                     GTVDDFLKLLNKEKSCQAVKDNDGGTINFTEKNDDKNNNNKDKGTFYRSEYCQVCPDC
                     GVKYNGSGWEEKKKNDQCNIKLYKPKKDAPHTPIKILKSGEGKEEIEKKLEAFCDKKD
                     GGNSDSSLYDPWQCYQFDQLEKDEKEEGVDDRNYDNDVRTGGGLCILQKKNGEENGKK
                     QKTYNDFFNFWVAHMLKDSIHWKKKLEKCLQNGTKTRCRNNEKCNKECECFQRWVEKK
                     KTEWGKIKDHFKTQNIGDETNCDPIVTLEGVLKLQFLNEDSTQDKQNSLDSEELKHLK
                     HLSEMLQETSGDGLTCGASDNEKETLMDKLIEHEEGIAKECLRKQNECEKKAKPEGRS
                     DSHDDPQPPDDADNEDDLDDEDDEDEEEEVQVEDNTQEEGEQPVVPQQEEGSSSPTPA
                     PAGPDVCDIVSQLFSDPSQFSDACTLKYVTGKNYGWKCIPSGNTSDTTGSESEATGAR
                     QRRDTDSSGDTTGGKDGATGGLCIPPRRRRLYVGGLSQWASQRTQGETSSQSGENLLE
                     AFIQSAAIETFFLWHKYKAENTKTQGVGAGGADFLPATSSVATALAPGAVPSRPSLQL
                     LSGVGVPGEPGMAPGVKSIPVPPLGVGVGGIPGVGALGGGALGPGGPVGLDGVPGQAQ
                     PLTLLRPGVLGNGLQSPQSRLRTLDGHFAGGESEDKTPQQWLQQGHIPPDFLRLMFYT
                     LADYRDICIGGDRDIVGDTIVSNTEGSSSSKIKISEKIKEILNHDNKQEPAPKPSVEK
                     TTPTEWWSQNGQHIWNGMICALTYTDSGGSITEDKDVRDKLIDKDTGKPQKNGDNDYT
                     YEKVELKDDESGPKGNDTIQPATLKDFVEIPTYFRWLHEWGSDFCGKRARMLKDVKDN
                     CRNIDKAGHHYCSGDGYDCTRDVIERNDKFVDLNCLGCYKQCRKYKKWIDIKFVEYHN
                     QEKKYKDEYGKLTKDKSSDDKKLEGYKCAENFLKELKHCKPSEDNNDQDNKINFDKPE
                     KTFNPSKYCKACPVYGVKYIGGNYIPNEEKDYKSKKGRVKKENDTIPKNIEVLVLGRK
                     GEEKDKDKHLHDACKNTGLFEVARYEQWNCQKKKGIDQCKITKFANDIDFDKDIVFNE
                     FFQRWLRYFVQDYNKLKDKINPCTKKETEKEKEKSYKCTQGCNDKCECVKEWLSKKKQ
                     EWTQIKTLYKQYSKISDQEIAFRVKSYFVDQGLFDNDYKKAQEVVEKPCDKEKLWGCT
                     GDNLKEGEDPGKCHMGDFITNLISKLQKKIDDCNKNQAQNSVETQPSDENTAQCQDTH
                     PDDEEDLLLEENENQVAQPNICPNQVEDKKIEEEVEKCETAQTTAEETAAAGGERQTP
                     PAPAPAAPPSPPRPLPKPKPPKPDLPPALKNAMLSSTIMWSVGIGFAAFTYFFLKKKT
                     KASVGNLFQILQIPKGDYDIPTLKSSNRYIPYASDRYKGKTYIYMEGDSSGDEKYAFM
                     SDTTDVTSSESEYEELDVNDIYVPGSPKYKTLIEVVLEPSGNNTPTSDIPSDNTPTPQ
                     PITDDEWNQLKHDFISNMLQNTQNTEPNILHDNVDNNTHPTMSRHNMDQKPFIMSIHD
                     RNLFSGEEYNYDMFNSGNNPINISDSTNSMDSLTSNNHSPYNDKNDLYSGIDLINDAL
                     SGNHIDIYDEMLKRKENELFGTQHHPKNITSNRVVTQTSSDDPIHNQLNLFHTWLDRH
                     RDMCEKLKNDNERLAKLKEEWENETHSGDINSGIPSGNHVLNTDVSIQIHMDNPKPIN
                     EFTNMDTSPDKSTMDTIIDDLEKYNEPYYYDFYEYDIYYDVNDDDKTSMDNNNNLVNK
                     NNPVDSNSSTYNHHNPADINKTFVDINNHNQHPIEKPTKIQIEMNSNNREVDEQQYPI
                     ADIWNI"
misc_feature         13697. .13898
                     /note="GC-rich region"
gene                 complement(join(17269. .18690,19611. .25169))
                     /gene="PFD0635c"
                     /note="synonym: VAR"
CDS                  complement(join(17269. .18690,19611. .25169))
                     /gene="PFD0635c"
                     /note="Similar to Plasmodium falciparum var, mal4p2.58
                     mal4p2.58, vaR SWALL:Q9U0G5 (EMBL:AL035475) (2277 aa)
                     fasta scores: E(): 1.1e-129, 49.19% id in 2498 aa, and to
                     Plasmodium falciparum variant-specific surface protein
                     var-7 SWALL:Q26034 (EMBL:L42636) (2182 aa) fasta scores:
                     E(): 7.9e-126, 46.25% id in 2441 aa"
                     /codon_start=1
                     /product="erythrocyte membrane protein 1 (PfEMP1)"
                     /protein_id="CAD49167.1"
                     /db_xref="GI:23498197"
                     /db_xref="GOA:Q8I1V6"
                     /db_xref="InterPro:IPR004258"
                     /db_xref="UniProt/TrEMBL:Q8I1V6"
                     /translation="MAPTSGGGGTKDESAKHALDSIGKRVHAQVQNEAKQRSNGDLKG
                     LLTSATLSGGESAFTENPCELIKEKREKLLGARGERNPCGNGSASEKRFSEVSGGECD
                     DKKIEGNGRNNGGACAPYRRLSLCNKNFQKINNYDSSKAKHNLLVDVCMAANYEAQSL
                     IPYHDQYDATYPGSDFSMCTMLARSFADIGDIIRGKDLYRGKKKKKQNGKETETERDQ
                     LESKLKEIFGDIYNELTNGRNGVKDHYQDTTDYFQLREDWWTANRHTVWKAITCGTHE
                     GDTYFRPTCSNRQGPSQAHHYCRCNGDKPDDDKPNTDPPTYFDYVPQYLRWFEEWAED
                     FCRKKKKYVDIVKKYCRGVYNDVPRYCSRNGYDCEQTIYKKGYFVIDKGCINCLYACN
                     PYVEWIEKQKEQFDKQKKKYTDEINEASRSSRRQKRGARSTGSSSNYDGYESKFYNIL
                     KDDYGTVDDFLKLLNKEKSCQAVKDNDGGTINFTEKNDDKNNNNKDKGTFYRSEYCQV
                     CPDCGVKYNGSGWEEKKKNDQCNIKLYKPKKDAPHTPIKILKSGEGKEEIEKKLEAFC
                     DKKDGGNSDSSLYDPWQCYQFDQLEKDEKEEGVDDRNYDNDVRTGGGLCILPNQKKNK
                     EEGANTSEKDHDEIQKTFNPFFYYWVAHMLKDSIHWKKKLEKCLQNGTKTRCRNNEKC
                     NKECECFQRWVEKKKTEWGKIKDHFKTQNIGDETNCDPIVTLEGVLKLQFLNEDSTQD
                     KQNSLDSEELKHLKHLSEMLQETSGDGLTCGASDNEKETLMDKLIEHEEGIAKECLRK
                     QNECEKKAKPEGRSDSHDDPQPPDDADNEDDLDDEDDEDEEEEVQVEDNTQEEGEQPV
                     VPQQEEGSSSPTPAPAGPDVCDIVSQLFSDPSQFSDACTLKYVTGKNYGWKCIPSGNT
                     SDTTGSESEATGARQRRDTDSSGDTTGGKDGATGGLCIPPRRRRLYVGGLSQWASQRT
                     QGETSSQSGENLLEAFIQSAAIETFFLWHKYKAENTKTQGVGAGGADFLPATSSVATA
                     LAPGAVPSRPSLQLLSGVGVPGEPGMAPGVKSIPVPPLGVGVGGIPGVGALGGGALGP
                     GGPVGLDGVPGQAQPLTLLRPGVLGNGLQSPQSRLRTLDGHFAGGESEDKTPQQWLQQ
                     GHIPPDFLRLMFYTLADYRDICIGGDRDIVGDTIVSNTEGSSSSKIKISEKIKEILNH
                     DNKQEPAPKPSVEKTTPTEWWSQNGQHIWNGMICALTYTDSGGSITEDKDVRDKLIDK
                     DTGKPQKNGDNDYTYEKVELKDDESGPKGNDTIQPATLKDFVEIPTYFRWLHEWGSDF
                     CGKRARMLKDVKDNCRNIDKAGHHYCSGDGYDCTRDVIERNDKFVDLNCLGCYKQCRK
                     YKKWIDIKFVEYHNQEKKYKDEYGKLTKDKSSDDKKLEGYKCAENFLKELKHCKPSED
```

```
                     NNDQDNKINFDKPEKTFNPSKYCKACPVYGVKYIGGNYIPNEEKDYKSKKGRVKKEND
                     TIPKNIEVLVLGRKGEEKDKDKHLHDACKNTGLFEVARYEQWNCQKKKGIDQCKITKF
                     ANDIDFDKDIVFNEFFQRWLRYFVQDYNKLKDKINPCTKKETEKEKEKSYKCTQGCND
                     KCECVKEWLSKKKQEWTQIKTLYKQYSKISDQEIAFRVKSYFVDQGLFDNDYKKAQEV
                     VEKPCDKEKLWGCTGDNLKEGEDPGKCHMGDFITNLISKLQKKIDDCNKNQAQNSVET
                     QPSDENTAQCQDTHPDDEEDLLLEENENQVAQPNICPNQVEDKKIEEEVEKCETAQTT
                     AEETAAAGGERQTPPAPAPAAPPSPPRPLPKPKPPKPDLPPALKNAMLSSTIMWSVGI
                     GFAAFTYFFLKKKTKASVGNLFQILQIPKGDYDIPTLKSSNRYIPYASDRYKGKTYIY
                     MEGDSSGDEKYAFMSDTTDVTSSESEYEELDVNDIYVPGSPKYKTLIEVVLEPSGNNT
                     PTSDIPSDNTPTPQPITDDEWNQLKHDFISNMLQNTQNTEPNILHDNVDNNTHPTMSR
                     HNMDQKPFIMSIHDRNLFSGEEYNYDMFNSGNNPINISDSTNSMDSLTSNNHSPYNDK
                     NDLYSGIDLINDALSGNHIDIYDEMLKRKENELFGTQHHPKNITSNRVVTQTSSDDPI
                     HNQLNLFHTWLDRHRDMCEKLKNDNERLAKLKEEWENETHSGDINSGIPSGNHVLNTD
                     VSIQIHMDNPKPINEFTNMDTSPDKSTMDTIIDDLEKYNEPYYYDFYEYDIYYDVNDD
                     DKTSMDNNNNLVNKNNPVDSNSSTYNHHNPADINKTFVDINNHNQHPIEKPTKIQIEM
                     NSNNREVDEQQYPIADIWNI"
     gene            complement(join(28486. .29547,29708. .29776))
                     /gene="PFD0640c"
                     /note="synonym: RIF"
     CDS             complement(join(28486. .29547,29708. .29776))
                     /gene="PFD0640c"
                     /note="Similar to Plasmodium falciparum rifin pfc1115w,
                     mal3p7.54 SWALL:O97311 (EMBL:AL034559) (371 aa) fasta
                     scores: E(): 4.8e-58, 49.73% id in 382 aa, and to
                     Plasmodium falciparum rifin pfb0035C SWALL:O96113
                     (EMBL:AE001367) (375 aa) fasta scores: E(): 1.9e-46,
                     43.52% id in 386 aa"
                     /codon_start=1
                     /product="RIFIN"
                     /protein_id="CAD49168.1"
                     /db_xref="GI:23498198"
                     /db_xref="InterPro:IPR002858"
                     /db_xref="InterPro:IPR006373"
                     /db_xref="UniProt/TrEMBL:Q8I1V5"
                     /translation="MKVHYISILLFSHPLNILLTSSQVYNQKNPYLTPHKTNPKSVKT
                     CRSLCECELYAPSNYDNDPEMKKVKQQFDDRTSQRSHEYHKRIQENRQKCKEQCEKDI
                     QKIILKDKIEKELTEKLGALQTEIRSDAIPTCTCQKSVGYKVEKTCLKCGGILGVGVA
                     PSLGLLGEIGGLVINNWKNTPFYEAFVAFAQKEGIAAGKIASDAARIDTVISGIISNF
                     EVHTINGSTLANAITLETLKDDTILTKALHFEYGSMCVNTPTDDKLICAYGMRAGLVQ
                     GKSASPEAVIRSSVKTLLKNADNVASQAAQTTANETTSGMIKAELSKIASAGANTYSA
                     ITYSVTAILVIVLVMVIIYLILRYRRKKKMKKKLQYIKLLKE"
     gene            join(34475. .34549,34788. .35753)
                     /gene="PFD0645w"
                     /note="synonym: RIF"
     CDS             join(34475. .34549,34788. .35753)
                     /gene="PFD0645w"
                     /note="Signal peptide predicted for mal1p4.07 by SignalP
                     2.0 HMM (Signal peptide probabilty 0.895, signal anchor
                     probability 0.008) with cleavage site probability 0.389
                     between residues 24 and 25
```

```
Query Match             2.5%;  Score 49.4;  DB 2;  Length 349751;
Best Local Similarity   45.2%;  Pred. No. 1.7;
Matches  182;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy       1309 AATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCA 1368
              ||   || | ||||  || |   |  |   | ||| |||   || ||| |  | ||   |
Db     302480 AAAAATAATACATATATAATAAAACATATGTTATATTTTAATCACTTTACATAAAAAATA 302421

Qy       1369 TCTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGA 1428
              |  |   | | ||     |     ||||  ||  | |   |||| ||   ||||   | |
Db     302420 TACAATTATATTAATAAATATATTAACATATGAATATATGAATATATGAATATTTGTACA 302361

Qy       1429 GCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAA 1488
                |    || || |   ||  |  ||| |||    | |   ||||   |      ||
Db     302360 TTTTGTTTTATTTATTTTATTTTATTTTTATTTTATTATTTTCTTTTTTCTTTAATTTCT 302301

Qy       1489 GGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATT 1548
              | | ||  |  | | | | | | |  | ||||  | | ||  |  |  ||| |  |||
Db     302300 ATGATTTTCAAAATTTTAAGAAATACATATTTTTACAAAAGAACAAAATTGTTATCAATG 302241

Qy       1549 CATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAG 1608
              |||  ||        | |   | || |  |   |||       |        ||  ||
```

```
Db     302240 GATTGTATATATATGGAGAATATAATTTAACACGTTCAAAAAAAAAAATAATAAAATAAT 302181

Qy       1609 TCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCT 1668
                |  |    |  ||  |||| | |||  |   || || |  || ||      |
Db     302180 ATAAAATAATAAAAAATAAAATAAAAATAAAAATAAAATAAAATAAATAAATAAATAA 302121

Qy       1669 CAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATA 1711
                ||  | |||||| | | |||  ||  || | |   | |||| |
Db     302120 CACAAAAAAAATGACATAAAACAAATATATAATTATAAAAAA 302078
```

```
RESULT 39
CR405684/c
LOCUS       CR405684              155357 bp    DNA     linear   HTG 21-MAR-2005
DEFINITION  Danio rerio clone DKEY-186I17, WORKING DRAFT SEQUENCE, 4 unordered
            pieces.
ACCESSION   CR405684
VERSION     CR405684.3  GI:61673441
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 155357)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 21, 2005 this sequence version replaced gi:47058833.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK186I17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 154173 bases at least Q40
            Consensus quality: 154392 bases at least Q30
            Consensus quality: 154589 bases at least Q20
            Insert size: 155057; sum-of-contigs
            Insert size: 160173; 1.4% error; agarose-fp
            Quality coverage: 8.48x in Q20 bases; sum-of-contigs Quality
            coverage: 8.21x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    39024: contig of 39024 bp in length
            *    39025    39124: gap of 100 bp
            *    39125    96064: contig of 56940 bp in length
            *    96065    96164: gap of 100 bp
            *    96165   109642: contig of 13478 bp in length
            *   109643   109742: gap of 100 bp
            *   109743   155357: contig of 45615 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155357
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-186I17"
                     /clone_lib="DanioKey"
     misc_feature    1. .39024
                     /note="assembly_fragment:00157
```

```
                     fragment_chain:1"
     misc_feature    39125. .96064
                     /note="assembly_fragment:01230
                     fragment_chain:1"
     misc_feature    96165. .109642
                     /note="assembly_fragment:00004"
     misc_feature    109743. .155357
                     /note="assembly_fragment:00650.0"
ORIGIN

  Query Match             2.5%;  Score 49.2;  DB 14;  Length 155357;
  Best Local Similarity   51.4%;  Pred. No. 2.1;
  Matches  114;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy       1535 ATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCAC 1594
             |||| || | |   |   || ||      | ||| |   || | |  || | | |    |
Db     110081 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 110022

Qy       1595 TTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAAC 1654
              |||    |   |   ||      ||  || |||||  | ||  ||| |||| | || ||
Db     110021 ATATATATATATATATATATATATAAAAAAAAAAAAAAAAAAAAAAATATATAATAAAAAAAAA 109962

Qy       1655 AAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTT 1714
              ||     |||  |    ||| ||| || ||||| |||  |  |  ||   |||| || |
Db     109961 AAAAAATATTTATTATTAATATAAAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAAATAT 109902

Qy       1715 GTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
                || ||    | ||| | ||   |  |||||  |  | |||
Db     109901 AATTTATTTTTTTATTAATAAAAAAAAAAAAAAAAAAAAT 109860


RESULT 40
CR354428/c
LOCUS       CR354428              214395 bp    DNA     linear   VRT 27-JUN-2005
DEFINITION  Zebrafish DNA sequence from clone DKEY-20D16 in linkage group 6,
            complete sequence.
ACCESSION   CR354428
VERSION     CR354428.16  GI:68262651
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 214395)
  AUTHORS   Auger,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jun 27, 2005 this sequence version replaced gi:66710852.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            DKEY-20D16 is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.
FEATURES             Location/Qualifiers
     source          1. .214395
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="6"
                     /clone="DKEY-20D16"
                     /clone_lib="DanioKey"
ORIGIN

  Query Match             2.4%;  Score 49;  DB 5;  Length 214395;
  Best Local Similarity   47.3%;  Pred. No. 2.2;
  Matches  148;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy       1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
             ||| ||| |  |||||   |||||  |  |       |  | || | || |  ||  |
Db       5245 ATGGAAAAACCTCATTTATTTGGCAAAGGTGATGCTAAATTTCAAGTTTTAGGCAAAAAA 5186

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
             || | ||  || | || |   ||  || |||||  | ||| ||| ||  || |   | |
Db       5185 ATAAAAATAAAATAATAAATAAAAAAAAAAAAAAAAAAAAATATATATATATATATATAT 5126

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
                | |   | ||||| | |||   | |   |   || |  |  |  | || ||  |
Db       5125 ATATATATATATATATATATATATATATATATATATATATATATATATATATATGTTATT 5066

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATC 1776
             ||  ||     |  || |    |  ||  | | |  |||   | | |  | || | |
Db       5065 TTTTTCTTTAAAGAGAAAGAAAAAAAAGTAACAGTGAGATTACATTTTAATAAAATTCTT 5006

Qy       1777 TTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCT 1836
             || |    |   | | | |||   | |    ||||      | |    ||||  |  ||
Db       5005 TTAATAAGACAAGTTAATAGTGTTTTCTGAAACCTGGAACTGCAAAAATGAAATAACCCT 4946

Qy       1837 TTGTCATAAAGTC 1849
                |||    | ||||
Db       4945 TTGAGCAACAGTC 4933


RESULT 41
AC087447/c
LOCUS       AC087447              113695 bp    DNA     linear   PRI 17-JUL-2002
DEFINITION  Homo sapiens chromosome 11, clone RP11-655D7, complete sequence.
ACCESSION   AC087447
VERSION     AC087447.5  GI:21886886
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 113695)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-655D7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113695)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
```

Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 113695)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (07-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 113695)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (17-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jul 17, 2002 this sequence version replaced gi:21702864.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L11851
Center clone name: 655_D_7

Only the last 113.7 kb of this clone are being submitted.
The remainder overlaps accession number AC010768 [WICGR project L2144].

FEATURES	Location/Qualifiers
source	1. .113695
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="11"
	/map="11"
	/clone="RP11-655D7"
	/clone_lib="RPCI-11 Human Male BAC"
repeat_region	30. .329
	/rpt_family="AluSq"
repeat_region	complement(432. .558)
	/rpt_family="MIR"
repeat_region	complement(1389. .1496)
	/rpt_family="MIR"
repeat_region	complement(1761. .1936)
	/rpt_family="L1ME3A"
repeat_region	1945. .1972
	/rpt_family="AT_rich"
repeat_region	complement(3265. .3466)
	/rpt_family="MIR"
repeat_region	4091. .4207
	/rpt_family="CT-rich"
repeat_region	complement(4216. .4525)
	/rpt_family="AluJb"
repeat_region	4645. .4752
	/rpt_family="MIR"
repeat_region	4754. .4861
	/rpt_family="(CATATA)n"
repeat_region	complement(4979. .5082)
	/rpt_family="MIR"
repeat_region	5779. .5845
	/rpt_family="MER113"
repeat_region	6272. .6302
	/rpt_family="(TTTA)n"
repeat_region	6303. .6326
	/rpt_family="AT_rich"
repeat_region	complement(6327. .11887)
	/rpt_family="L1PA7"
repeat_region	complement(11889. .12180)
	/rpt_family="AluSx"
repeat_region	complement(12994. .13169)
	/rpt_family="MIR"
repeat_region	complement(13271. .13368)
	/rpt_family="L2"
repeat_region	13389. .13786
	/rpt_family="MLT1J"
repeat_region	13867. .14161
	/rpt_family="AluSx"
repeat_region	14265. .14569
	/rpt_family="AluSc"
repeat_region	complement(14598. .14709)
	/rpt_family="MIR"
repeat_region	complement(14918. .15517)
	/rpt_family="L2"
repeat_region	complement(15678. .15736)
	/rpt_family="MIR3"
repeat_region	16898. .17273
	/rpt_family="MLT1B"
repeat_region	17551. .18001
	/rpt_family="L1MC4a"
repeat_region	18011. .18187
	/rpt_family="L1MC4"
repeat_region	18382. .18509
	/rpt_family="MIR"
repeat_region	complement(18517. .18824)
	/rpt_family="AluJb"
repeat_region	19081. .19157

```
                     /rpt_family="MIR"
     repeat_region   complement(19165. .19335)
                     /rpt_family="MLT1I"
     repeat_region   complement(19362. .19484)
                     /rpt_family="MLT1I"
     repeat_region   19489. .19645
                     /rpt_family="MIR"
     repeat_region   19850. .19883
                     /rpt_family="(TTTTTG)n"
     repeat_region   complement(19962. .20006)
                     /rpt_family="L2"
     repeat_region   20913. .21209
                     /rpt_family="AluSp"
     repeat_region   21213. .21318
                     /rpt_family="MIR"
     repeat_region   complement(21562. .21731)
                     /rpt_family="L2"
     repeat_region   21987. .22134
                     /rpt_family="L3"
     repeat_region   23078. .23145
                     /rpt_family="A-rich"
     repeat_region   complement(23579. .23766)
                     /rpt_family="MIR"
     repeat_region   complement(23824. .23974)
                     /rpt_family="FRAM"
     repeat_region   complement(23994. .24149)
                     /rpt_family="MIR"
     repeat_region   complement(24468. .24524)
                     /rpt_family="L1MB4"
     repeat_region   complement(24686. .24925)
                     /rpt_family="MER20"
     repeat_region   24948. .25387

  Query Match             2.4%;  Score 48.4;  DB 8;  Length 113695;
  Best Local Similarity   71.1%;  Pred. No. 3.3;
  Matches   64;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         50 GTGCATATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAA 109
              | |||   || |     |  | | ||||||||| |||||||  |||  ||||   | |||
Db      79588 GAGCAGGGAGGTATTCCCTCTTGGTTGCAAGTGGCAGAAACCTAATGAAAACTGGCTTAA 79529

Qy        110 GTCAAAAGGAAAATATATTGGCTCATGTAA 139
              |  |||||||||||||||||||| |||||
Db      79528 GCAAAAAGGAAAATATATTGGCTCCTGTAA 79499


RESULT 42
AC021170/c
LOCUS       AC021170               176697 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-655D7, WORKING DRAFT
            SEQUENCE, 24 unordered pieces.
ACCESSION   AC021170
VERSION     AC021170.4  GI:7235310
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 176697)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176697)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 13, 2000 this sequence version replaced gi:7024164.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0655D07
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 164861 bases at least Q40
            Consensus quality: 168822 bases at least Q30
            Consensus quality: 170901 bases at least Q20
            Insert size: 194000; agarose-fp
            Insert size: 174397; sum-of-contigs
            Quality coverage: 3.57 in Q20 bases; agarose-fp
            Quality coverage: 4.02 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1086: contig of 1086 bp in length
            *     1087     1186: gap of unknown length
            *     1187     4671: contig of 3485 bp in length
            *     4672     4771: gap of unknown length
            *     4772     7902: contig of 3131 bp in length
            *     7903     8002: gap of unknown length
            *     8003    11020: contig of 3018 bp in length
            *    11021    11120: gap of unknown length
            *    11121    14825: contig of 3705 bp in length
            *    14826    14925: gap of unknown length
            *    14926    17464: contig of 2539 bp in length
            *    17465    17564: gap of unknown length
            *    17565    21883: contig of 4319 bp in length
            *    21884    21983: gap of unknown length
            *    21984    25335: contig of 3352 bp in length
            *    25336    25435: gap of unknown length
            *    25436    28938: contig of 3503 bp in length
            *    28939    29038: gap of unknown length
            *    29039    32405: contig of 3367 bp in length
            *    32406    32505: gap of unknown length
            *    32506    36284: contig of 3779 bp in length
            *    36285    36384: gap of unknown length
            *    36385    39520: contig of 3136 bp in length
            *    39521    39620: gap of unknown length
            *    39621    43560: contig of 3940 bp in length
            *    43561    43660: gap of unknown length
            *    43661    48581: contig of 4921 bp in length
            *    48582    48681: gap of unknown length
            *    48682    52884: contig of 4203 bp in length
            *    52885    52984: gap of unknown length
            *    52985    57434: contig of 4450 bp in length
            *    57435    57534: gap of unknown length
            *    57535    62596: contig of 5062 bp in length
            *    62597    62696: gap of unknown length
            *    62697    69408: contig of 6712 bp in length
            *    69409    69508: gap of unknown length
            *    69509    76036: contig of 6528 bp in length
            *    76037    76136: gap of unknown length
            *    76137    83650: contig of 7514 bp in length
            *    83651    83750: gap of unknown length
            *    83751    93283: contig of 9533 bp in length
            *    93284    93383: gap of unknown length
            *    93384   102351: contig of 8968 bp in length
            *   102352   102451: gap of unknown length
            *   102452   129254: contig of 26803 bp in length
            *   129255   129354: gap of unknown length
            *   129355   176697: contig of 47343 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .176697
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-655D7"
     gap             1087. .1186
                     /estimated_length=unknown
     gap             4672. .4771
                     /estimated_length=unknown
     gap             7903. .8002
                     /estimated_length=unknown
     gap             11021. .11120
                     /estimated_length=unknown
     gap             14826. .14925
                     /estimated_length=unknown
     gap             17465. .17564
                     /estimated_length=unknown
     gap             21884. .21983
                     /estimated_length=unknown
     gap             25336. .25435
                     /estimated_length=unknown
     gap             28939. .29038
                     /estimated_length=unknown
     gap             32406. .32505
                     /estimated_length=unknown
     gap             36285. .36384
                     /estimated_length=unknown
     gap             39521. .39620
                     /estimated_length=unknown
     gap             43561. .43660
                     /estimated_length=unknown
     gap             48582. .48681
                     /estimated_length=unknown
     gap             52885. .52984
                     /estimated_length=unknown
     gap             57435. .57534
                     /estimated_length=unknown
     gap             62597. .62696
                     /estimated_length=unknown
     gap             69409. .69508
                     /estimated_length=unknown
     gap             76037. .76136
                     /estimated_length=unknown
     gap             83651. .83750
                     /estimated_length=unknown
     gap             93284. .93383
                     /estimated_length=unknown
     gap             102352. .102451
                     /estimated_length=unknown
     gap             129255. .129354
                     /estimated_length=unknown
ORIGIN

  Query Match              2.4%;  Score 48.4;  DB 14;  Length 176697;
  Best Local Similarity   71.1%;  Pred. No. 3.1;
  Matches   64;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          50 GTGCATATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAA 109
               | |||   || |     |  | | ||||||||| |||||||  |||  ||||   | |||
Db      154947 GAGCAGGGAGGTATTCCCTCTTGGTTGCAAGTGGCAGAAACCTAATGAAAACTGGCTTAA 154888

Qy         110 GTCAAAAGGAAAATATATTGGCTCATGTAA 139
               |  ||||||||||||||||||||| |||||
Db      154887 GCAAAAAGGAAAATATATTGGCTCCTGTAA 154858


RESULT 43
AC019143/c
LOCUS       AC019143               190737 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-706A13, WORKING DRAFT
            SEQUENCE, 28 unordered pieces.
ACCESSION   AC019143
VERSION     AC019143.3  GI:7715665
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 190737)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190737)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 7, 2000 this sequence version replaced gi:7024180.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0706A13
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 78%
            Sequencing vector: plasmid; 22%
            Chemistry: Dye-primer ET; 78% of reads
            Chemistry: Dye-terminator Big Dye; 22% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 167325 bases at least Q40
            Consensus quality: 175317 bases at least Q30
            Consensus quality: 180039 bases at least Q20
            Insert size: 193000; agarose-fp
            Insert size: 188037; sum-of-contigs
            Quality coverage: 3.04 in Q20 bases; agarose-fp
            Quality coverage: 3.18 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1325: contig of 1325 bp in length
            *    1326     1425: gap of unknown length
            *    1426     2586: contig of 1161 bp in length
            *    2587     2686: gap of unknown length
            *    2687     4338: contig of 1652 bp in length
            *    4339     4438: gap of unknown length
            *    4439     5989: contig of 1551 bp in length
            *    5990     6089: gap of unknown length
            *    6090     8269: contig of 2180 bp in length
            *    8270     8369: gap of unknown length
            *    8370    10239: contig of 1870 bp in length
            *   10240    10339: gap of unknown length
            *   10340    13484: contig of 3145 bp in length
            *   13485    13584: gap of unknown length
            *   13585    18186: contig of 4602 bp in length
            *   18187    18286: gap of unknown length
            *   18287    23764: contig of 5478 bp in length
            *   23765    23864: gap of unknown length
            *   23865    27992: contig of 4128 bp in length
            *   27993    28092: gap of unknown length
            *   28093    31755: contig of 3663 bp in length
            *   31756    31855: gap of unknown length
            *   31856    36229: contig of 4374 bp in length
            *   36230    36329: gap of unknown length
            *   36330    40157: contig of 3828 bp in length
```

```
            *    40158    40257: gap of unknown length
            *    40258    45494: contig of 5237 bp in length
            *    45495    45594: gap of unknown length
            *    45595    51920: contig of 6326 bp in length
            *    51921    52020: gap of unknown length
            *    52021    58426: contig of 6406 bp in length
            *    58427    58526: gap of unknown length
            *    58527    64892: contig of 6366 bp in length
            *    64893    64992: gap of unknown length
            *    64993    72930: contig of 7938 bp in length
            *    72931    73030: gap of unknown length
            *    73031    81807: contig of 8777 bp in length
            *    81808    81907: gap of unknown length
            *    81908    89406: contig of 7499 bp in length
            *    89407    89506: gap of unknown length
            *    89507    99182: contig of 9676 bp in length
            *    99183    99282: gap of unknown length
            *    99283   111834: contig of 12552 bp in length
            *   111835   111934: gap of unknown length
            *   111935   126258: contig of 14324 bp in length
            *   126259   126358: gap of unknown length
            *   126359   134920: contig of 8562 bp in length
            *   134921   135020: gap of unknown length
            *   135021   145335: contig of 10315 bp in length
            *   145336   145435: gap of unknown length
            *   145436   162083: contig of 16648 bp in length
            *   162084   162183: gap of unknown length
            *   162184   175617: contig of 13434 bp in length
            *   175618   175717: gap of unknown length
            *   175718   190737: contig of 15020 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190737
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-706A13"
     misc_feature    1. .1325
                     /note="assembly_name:Contig4"
     gap             1326. .1425
                     /estimated_length=unknown
     misc_feature    1426. .2586
                     /note="assembly_name:Contig5"
     gap             2587. .2686
                     /estimated_length=unknown
     misc_feature    2687. .4338
                     /note="assembly_name:Contig6"
     gap             4339. .4438
                     /estimated_length=unknown
     misc_feature    4439. .5989
                     /note="assembly_name:Contig7"
     gap             5990. .6089
                     /estimated_length=unknown
     misc_feature    6090. .8269
                     /note="assembly_name:Contig8"
     gap             8270. .8369
                     /estimated_length=unknown
     misc_feature    8370. .10239
                     /note="assembly_name:Contig9"
     gap             10240. .10339
                     /estimated_length=unknown
     misc_feature    10340. .13484
                     /note="assembly_name:Contig10"
     gap             13485. .13584
                     /estimated_length=unknown
     misc_feature    13585. .18186
                     /note="assembly_name:Contig11"
     gap             18187. .18286
                     /estimated_length=unknown
     misc_feature    18287. .23764
                     /note="assembly_name:Contig12"
     gap             23765. .23864
                     /estimated_length=unknown
     misc_feature    23865. .27992
                     /note="assembly_name:Contig13"
     gap             27993. .28092
                     /estimated_length=unknown
     misc_feature    28093. .31755
                     /note="assembly_name:Contig14"
     gap             31756. .31855
                     /estimated_length=unknown
     misc_feature    31856. .36229
                     /note="assembly_name:Contig15"
     gap             36230. .36329
                     /estimated_length=unknown
     misc_feature    36330. .40157
                     /note="assembly_name:Contig16"
     gap             40158. .40257
                     /estimated_length=unknown
     misc_feature    40258. .45494
                     /note="assembly_name:Contig17"
     gap             45495. .45594
                     /estimated_length=unknown
     misc_feature    45595. .51920
                     /note="assembly_name:Contig18"
     gap             51921. .52020
                     /estimated_length=unknown
     misc_feature    52021. .58426
                     /note="assembly_name:Contig19"
     gap             58427. .58526
                     /estimated_length=unknown
     misc_feature    58527. .64892
                     /note="assembly_name:Contig21"
     gap             64893. .64992
                     /estimated_length=unknown
     misc_feature    64993. .72930
                     /note="assembly_name:Contig22"
     gap             72931. .73030
                     /estimated_length=unknown
     misc_feature    73031. .81807
                     /note="assembly_name:Contig23"
     gap             81808. .81907
                     /estimated_length=unknown
     misc_feature    81908. .89406
                     /note="assembly_name:Contig24
                     clone_end:SP6
                     vector_side:right"
     gap             89407. .89506
                     /estimated_length=unknown
     misc_feature    89507. .99182
                     /note="assembly_name:Contig25"
     gap             99183. .99282
                     /estimated_length=unknown
     misc_feature    99283. .111834

  Query Match             2.4%;  Score 48.4;  DB 14;  Length 190737;
  Best Local Similarity   71.1%;  Pred. No. 3.1;
  Matches   64;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         50 GTGCATATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAA 109
              | |||   || |     |  | | ||||||||| |||||||  |||  ||||   | |||
Db     128302 GAGCAGGGAGGTATTCCCTCTTGGTTGCAAGTGGCAGAAACCTAATGAAAACTGGCTTAA 128243

Qy        110 GTCAAAAGGAAAATATATTGGCTCATGTAA 139
              |  |||||||||||||||||||| |||||
Db     128242 GCAAAAAGGAAAATATATTGGCTCCTGTAA 128213


RESULT 44
AC090163/c
LOCUS       AC090163              194642 bp    DNA     linear   PRI 30-AUG-2002
DEFINITION  Homo sapiens chromosome 11, clone RP11-706A13, complete sequence.
ACCESSION   AC090163
VERSION     AC090163.10  GI:22549876
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 194642)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-706A13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194642)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 194642)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 194642)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 30, 2002 this sequence version replaced gi:22123058.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11858
                Center clone name: 706_A_13
            --------------.
FEATURES             Location/Qualifiers
     source          1. .194642
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-706A13"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   2. .518
                     /rpt_family="MLT1E"
     repeat_region   783. .1081
                     /rpt_family="MER2"
     repeat_region   1224. .1372
                     /rpt_family="MIR"
     repeat_region   complement(1808. .2067)
                     /rpt_family="MLT1K"
     repeat_region   2222. .2587
                     /rpt_family="MLT1L"
     repeat_region   3040. .3139
                     /rpt_family="MIR"
     repeat_region   3187. .3281
                     /rpt_family="MIR"
     repeat_region   3879. .4029
                     /rpt_family="CT-rich"
     repeat_region   4069. .4219
                     /rpt_family="CT-rich"
     repeat_region   complement(4561. .4658)
                     /rpt_family="MIR3"
     repeat_region   4647. .4712
                     /rpt_family="MIR"
     repeat_region   complement(5523. .5629)
                     /rpt_family="L2"
     repeat_region   complement(6301. .6380)
                     /rpt_family="MIR"
     repeat_region   6391. .6529
                     /rpt_family="MIR3"
     repeat_region   6737. .6764
                     /rpt_family="(TTCA)n"
     repeat_region   7797. .7929
                     /rpt_family="L2"
     repeat_region   8080. .8206
                     /rpt_family="L2"
     repeat_region   complement(8240. .8446)
                     /rpt_family="MIR"
     repeat_region   complement(9198. .9353)
                     /rpt_family="L2"
     repeat_region   9534. .10011
                     /rpt_family="L1ME1"
     repeat_region   10033. .10064
                     /rpt_family="MIR"
     repeat_region   complement(10748. .10863)
                     /rpt_family="MIR"
     repeat_region   10936. .11071
                     /rpt_family="MIR3"
     repeat_region   complement(11078. .11628)
```

```
                     /rpt_family="MLT1E"
     repeat_region   12068. .12342
                     /rpt_family="AluSx"
     repeat_region   13226. .13323
                     /rpt_family="L2"
     repeat_region   13330. .13385
                     /rpt_family="MER103"
     repeat_region   13386. .13691
                     /rpt_family="AluSp"
     repeat_region   13692. .13805
                     /rpt_family="MER103"
     repeat_region   complement(13982. .14245)
                     /rpt_family="MLT1K"
     repeat_region   complement(14497. .14623)
                     /rpt_family="MIR"
     repeat_region   14996. .15197
                     /rpt_family="MIR"
     repeat_region   complement(15212. .15349)
                     /rpt_family="AluJo"
     repeat_region   15350. .15372
                     /rpt_family="(TTTAA)n"
     repeat_region   complement(15373. .15496)
                     /rpt_family="AluJo"
     repeat_region   16740. .16797
                     /rpt_family="MLT1B"
     repeat_region   16960. .17271
                     /rpt_family="AluSx"
     repeat_region   17306. .17419
                     /rpt_family="AluJo"
     repeat_region   17437. .19219
                     /rpt_family="L1PA5"
     repeat_region   complement(19220. .19523)
                     /rpt_family="AluY"
     repeat_region   19524. .23852
                     /rpt_family="L1PA5"
     repeat_region   23853. .24038
                     /rpt_family="AluJo"
     repeat_region   24044. .24596
                     /rpt_family="MLT1E"
     repeat_region   24598. .24635
                     /rpt_family="L2"
     repeat_region   24780. .24908
                     /rpt_family="L2"
     repeat_region   25217. .25434
                     /rpt_family="HAL1"
     repeat_region   25479. .26613
                     /rpt_family="HAL1"
     repeat_region   26614. .26822

  Query Match             2.4%;  Score 48.4;  DB 8;  Length 194642;
  Best Local Similarity   71.1%;  Pred. No. 3.1;
  Matches   64;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         50 GTGCATATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAA 109
              | |||   || |     |  | | ||||||||| |||||||  |||  ||||   | |||
Db      44544 GAGCAGGGAGGTATTCCCTCTTGGTTGCAAGTGGCAGAAACCTAATGAAAACTGGCTTAA 44485

Qy        110 GTCAAAAGGAAAATATATTGGCTCATGTAA 139
              |  ||||||||||||||||||||| |||||
Db      44484 GCAAAAAGGAAAATATATTGGCTCCTGTAA 44455


RESULT 45
AF267226
LOCUS       AF267226               1818 bp    DNA     linear   BCT 15-APR-2003
DEFINITION  Candidatus Carsonella ruddii natural-host Trioza urticae ATP
            synthase alpha subunit (atpA) gene, partial cds; ATP synthase gamma
            subunit (atpG) gene, complete cds; and ATP synthase beta subunit
            (atpD) gene, partial cds.
ACCESSION   AF267226
VERSION     AF267226.1  GI:14193360
KEYWORDS    .
SOURCE      Candidatus Carsonella ruddii
  ORGANISM  Candidatus Carsonella ruddii
            Bacteria; Proteobacteria; Gammaproteobacteria; Candidatus
            Carsonella.
REFERENCE   1  (bases 1 to 1818)
  AUTHORS   Thao,M.L., Clark,M.A., Burckhardt,D.H., Moran,N.A. and Baumann,P.
  TITLE     Phylogenetic analysis of vertically transmitted psyllid
            endosymbionts (Candidatus Carsonella ruddii) based on atpAGD and
            rpoC: comparisons with 16S-23S rDNA-derived phylogeny
  JOURNAL   Curr. Microbiol. 42 (6), 419-421 (2001)
   PUBMED   11381334
REFERENCE   2  (bases 1 to 1818)
  AUTHORS   Thao,M.L., Clark,M.A., Burckhardt,D.H., Moran,N.A. and Baumann,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2000) Section of Microbiology, University of
            California, Davis, One Shields Avenue, Davis, CA 95616, USA
FEATURES             Location/Qualifiers
     source          1. .1818
                     /organism="Candidatus Carsonella ruddii"
                     /mol_type="genomic DNA"
                     /specific_host="Trioza urticae"
                     /db_xref="taxon:114186"
     gene            <1. .656
                     /gene="atpA"
     CDS             <1. .656
                     /gene="atpA"
                     /codon_start=3
                     /transl_table=11
                     /product="ATP synthase alpha subunit"
                     /protein_id="AAK55931.1"
                     /db_xref="GI:14193361"
                     /translation="VFYLHSRLLERSSKVNKFFLEKKNIFNKNNGTLTAFPIIETLEG
                     DVTSFIPTNVISITDGQIFLDTNLFNSGIRPAINVGLSVSRVGGSAQYKIVKKLSGDI
                     RIMLAQYRELESFSKFSSDLDIETKNQLYNGEKITLLMKQKPNYNYSIIELVLILLII
                     KNVYFKKIPVKEIEFFEFNIINYIRNMKIEKIIKIDNINLEYYLIEIINFFIFGSLL"
     gene            653. .1405
                     /gene="atpG"
     CDS             653. .1405
                     /gene="atpG"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP synthase gamma subunit"
                     /protein_id="AAK55932.1"
                     /db_xref="GI:14193362"
                     /translation="MNIRDIKCKINILFNINKLTNTMSTISFSKMKKILKQCFILNKL
                     YLETKKIIFEIYNFKKNNFFCCILLTTNKGFCSNINSEIIKYCIKFLKTNKNLDLIII
                     GKKAVDFFNKKNIIITKKIIFDEKKDIFFSNEILKILNNYENVFFLSSKIFNNNIKII
                     KTNLYEKIKKNFYEIKNINIYKLINNYINFTLKYLYLENYFSELKLRMITMKSATDSS
                     KKIIKNMNTIKNKIRQFKVTQEMLEIINSINL"
     gene            1402. .>1818
                     /gene="atpD"
     CDS             1402. .>1818
                     /gene="atpD"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP synthase beta subunit"
                     /protein_id="AAK55933.1"
                     /db_xref="GI:14193363"
                     /translation="MTGRIIQILGSILDVEFKKKNIPKIYNALFIPEVNIYLEVQQQI
                     GKNIVRTIAFGNTNGMKRNSIVIDTGKPIMTPVGDCTLGRILNVLGEPIDNKGNLNFI
                     KKKPIHLSPPKFSEQIFNNNILETGIKIIDLLCPFLK"
ORIGIN

  Query Match             2.4%;  Score 48.2;  DB 1;  Length 1818;
  Best Local Similarity   45.9%;  Pred. No. 6;
  Matches  164;  Conservative    0;  Mismatches  193;  Indels    0;  Gaps    0;

Qy       1415 TATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTG 1474
              |||  ||||| | |  |||   | ||||    | |    ||| |   | | |   | |||
Db        472 TATTAATTATAAAAAATGTTTATTTTAAAAAAATACCAGTTAAAGAAATTGAATTTTTTG 531

Qy       1475 GCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTC 1534
                ||  ||||||        | || ||| ||| |    |      || |  | |    |
Db        532 AATTTAATATAATTAATTATATTAGAAATATGAAAATTGAAAAAATTATAAAAATTGATA 591
```

```
Qy       1535 ATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCAC 1594
Db        592 ATATAAATTTAGAATATTATTTAATAGAAATTATTAATTTTTTTATATTTGGTAGTTTAT 651

Qy       1595 TTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAAC 1654
Db        652 TATGAATATAAGAGATATTAAATGTAAAATTAATATTTTATTTAATATAAATAAATTAAC 711

Qy       1655 AAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTT 1714
Db        712 AAATACAATGAGTACGATATCGTTTTCTAAAATGAAAAAAATTTTAAAACAATGTTTCAT 771

Qy       1715 GTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAA 1771
Db        772 ATTAAATAAATTATATTTAGAAACAAAAAAAATAATTTTTGAAATTTATAATTTTAA 828
```

Search completed: January 22, 2006, 13:37:42
Job time : 6826.68 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2006, 09:21:22 ; Search time 805.081 Seconds
(without alignments)
16564.866 Million cell updates/sec

Title: US-09-728-552A-3_COPY_78595_80595
Perfect score: 2001
Sequence: 1 atctttatttgtatgagaaa.....aagtcattctgccagaattc 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	80595	2	AAV83939	Aav83939 HC-contig
	2	1984.8	99.2	18443	2	AAV83942	Aav83942 Bacterial
	3	1895.2	94.7	80240	2	AAV83940	Aav83940 NC-contig
	4	97.6	4.9	300	2	AAV83958	Aav83958 Bacterial
	5	58.8	2.9	2000	8	ADA71938	Ada71938 Rice gene
	6	52	2.6	2000	11	ACL37108	Acl37108 Rice stre
	7	49.4	2.5	431	3	AAC15209	Aac15209 Human sec
c	8	49.4	2.5	1428	14	ADZ70899	Adz70899 Human mat
	9	45	2.2	1428	14	ADZ70899	Adz70899 Human mat
	10	44.8	2.2	9238	6	ABK28365	Abk28365 DNA trans
c	11	44.4	2.2	3100	2	AAV20419	Aav20419 Human dis
c	12	44.4	2.2	3100	2	AAX86725	DNA seque
c	13	44.4	2.2	3100	9	ACC85073	Acc85073 Human MBC
c	14	44.4	2.2	3196	2	AAV20420	Aav20420 Human dis
c	15	44.4	2.2	5147	5	ABV25006	Abv25006 Human pro
	16	44.2	2.2	2000	8	ADA73182	Ada73182 Rice gene
	17	44.2	2.2	2000	11	ACL35363	Acl35363 Rice stre
c	18	44.2	2.2	82938	6	ABV72623	Abv72623 Human tra
	19	44	2.2	3891	6	ABX13938	Abx13938 Clostridi
	20	44	2.2	3891	6	AAD42516	Aad42516 Clostridi
c	21	44	2.2	7093	6	ABZ35416	Abz35416 Human gen
c	22	44	2.2	10126	4	AAI99260	Aai99260 Human exc
c	23	44	2.2	10126	5	AAI63610	Aai63610 Human kid
c	24	43.6	2.2	5072	13	ADR24206	Adr24206 Breast ca
c	25	43.6	2.2	5073	9	ACC85075	Acc85075 Human MBC
c	26	43.6	2.2	10369	4	AAS46304	Aas46304 Tumour su
c	27	43.6	2.2	10369	6	ABL32393	Abl32393 Human imm
c	28	43.4	2.2	1212	13	ADR85280	Adr85280 Aspergill
c	29	43.4	2.2	1236	14	ADZ70960	Adz70960 Human chr
c	30	43.4	2.2	3212	13	ADR84693	Adr84693 Aspergill
c	31	43	2.1	32216	4	AAL06925	Aal06925 Human rep
c	32	43	2.1	32216	4	ABA07741	Aba07741 Human ova
	33	43	2.1	243428	12	ADF51132	Adf51132 Human P-R
c	34	42.8	2.1	9963	6	ABL32695	Abl32695 Human imm
c	35	42.8	2.1	110000	14	AEB42401_32	Continuation (33 o
c	36	42.6	2.1	16724	6	ABL33091	Abl33091 Human imm
c	37	42.6	2.1	16724	6	ABL34537	Abl34537 Human met
c	38	42.6	2.1	16724	6	ABL70260	Abl70260 Chemicall
c	39	42.6	2.1	16724	7	ADS99798	Ads99798 Complemen
	40	42.4	2.1	416	8	ABX39419	Abx39419 Bovine ES
	41	42.4	2.1	1317	3	AAZ87220	Aaz87220 DNA encod
	42	42.4	2.1	1546	2	AAV30575	Aav30575 Clostridi
	43	42.4	2.1	1917	3	AAZ87219	Aaz87219 DNA encod
c	44	42.4	2.1	2000	8	ADA71938	Ada71938 Rice gene
	45	42.4	2.1	2532	3	AAZ87218	Aaz87218 DNA encod
	46	42.4	2.1	3891	2	AAT29244	Aat29244 C. botuli
	47	42.4	2.1	3891	14	ADZ36017	Adz36017 C. botuli
	48	42.4	2.1	4835	3	AAC64582	Aac64582 BoNT/A ne
	49	42.4	2.1	4835	14	ADW24419	Adw24419 Clostridi
	50	42.4	2.1	4835	14	ADZ60272	Adz60272 Clostridi
c	51	42.4	2.1	6881	6	ABL33380	Abl33380 Human imm
c	52	42.4	2.1	6881	6	ABK28249	Abk28249 DNA trans
c	53	42.2	2.1	6791	13	ADT71610	Adt71610 Protease-
c	54	42.2	2.1	6833	13	ADT71609	Adt71609 Protease-
c	55	42.2	2.1	89213	13	ABD33614	Abd33614 Murine ca
	56	42	2.1	520	13	ACN52918	Acn52918 Cotton an
c	57	42	2.1	6046	4	AAS45311	Aas45311 Chemicall
c	58	42	2.1	6046	6	ABK28150	Abk28150 DNA trans
c	59	42	2.1	29993	10	ADB37662	Adb37662 Human che
c	60	42	2.1	29993	10	ADB37660	Adb37660 Human che
c	61	42	2.1	38342	4	AAS46745	Aas46745 Tumour su
c	62	42	2.1	38342	6	ABK31506	Abk31506 Signal tr
	63	41.8	2.1	802	13	ADR64347	Adr64347 Cotton cD
	64	41.8	2.1	2672	12	ADQ23221	Adq23221 Human sof
c	65	41.8	2.1	9157	6	ABL33500	Abl33500 Human imm
c	66	41.6	2.1	13574	6	ABL33317	Abl33317 Human imm
c	67	41.6	2.1	68815	14	ADZ12783	Adz12783 Human can
	68	41.4	2.1	1350	14	ADZ71037	Adz71037 Human chr
c	69	41.4	2.1	2000	11	ACL37108	Acl37108 Rice stre
c	70	41.4	2.1	6220	6	ABL33300	Abl33300 Human imm
c	71	41.4	2.1	6223	6	AAS61177	Aas61177 Human gen
	72	41.2	2.1	424	8	ABX46053	Abx46053 Bovine ES
	73	41.2	2.1	2196	6	ABK65323	Abk65323 Arabidops
	74	41.2	2.1	2353	12	ADI43598	Adi43598 Plant tra
	75	41.2	2.1	2353	12	ADO61678	Ado61678 Transcrip
c	76	41.2	2.1	3312	2	AAQ79326	Aaq79326 Human ARD
c	77	41.2	2.1	3334	6	ABZ34989	Abz34989 Human gen
c	78	41.2	2.1	3334	13	ADR24783	Adr24783 Breast ca
c	79	41.2	2.1	5571	4	AAS46484	Aas46484 Tumour su
c	80	41.2	2.1	5571	6	ABK31347	Abk31347 Signal tr
c	81	41.2	2.1	5571	6	ABL70308	Abl70308 Chemicall
c	82	41.2	2.1	5571	6	AAS61242	Aas61242 Human gen
	83	41	2.0	895	13	ADT17682	Adt17682 Plant cDN
	84	41	2.0	8372	4	AAS46353	Aas46353 Tumour su
	85	41	2.0	42488	4	AAK66772	Aak66772 Human imm
c	86	40.8	2.0	321	10	ADD32808	Add32808 Human mit
c	87	40.8	2.0	321	10	ABX74672	Abx74672 Human cDN
	88	40.6	2.0	2000	11	ACL35887	Acl35887 Rice stre
c	89	40.6	2.0	2000	11	ACL35887	Acl35887 Rice stre
c	90	40.6	2.0	5922	6	ABL32450	Abl32450 Human imm
	91	40.6	2.0	6203	4	AAS45475	Aas45475 Chemicall
	92	40.6	2.0	6203	6	ABK28399	Abk28399 DNA trans

c	93	40.6	2.0	6794	6	ABK31265	Abk31265 Signal tr
c	94	40.6	2.0	6794	6	ABL70220	Abl70220 Chemicall
c	95	40.6	2.0	6794	6	AAS61175	Aas61175 Human gen
	96	40.6	2.0	24939	6	ABL70569	Abl70569 Chemicall
c	97	40.6	2.0	110000	6	ABA92787_0	Aba92787 Buchnera
c	98	40.4	2.0	15373	6	ABL32467	Abl32467 Human imm
	99	40.4	2.0	57130	8	ABQ77243	Abq77243 Human MAR
c	100	40.2	2.0	433	14	ADZ70956	Adz70956 Human chr

ALIGNMENTS

RESULT 1
AAV83939
ID AAV83939 standard; DNA; 80595 BP.
XX
AC AAV83939;
XX
DT 03-MAR-1999 (first entry)
XX
DE HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 8; Fig 6; 540pp; English.
XX
CC The present sequence represents the HC-contig derived from normal human
CC chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC produce an unusual chromosomal marker referred to as mardel(10). The
CC mardel(10) marker is mitotically stable and contains a functional
CC neocentromere at a location regarded as non-centromeric. This
CC neocentromere maps to q25.2 on chromosome 10. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in a
CC compatible cell, of replicating, acting as extra-chromosomal element and
CC segregating during cell division. The sequences can be used to construct
CC artificial chromosomes for use in gene therapy comprising a replicable,
CC segregating nucleic acid that confers a specific phenotype on cells.
CC Human artificial chromosomes can propagate in human cells and carry large
CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC they are not mutagenic. The artificial chromosomes are also useful for
CC generation of transgenic plants and animals, in production of proteins
CC and to make diagnostic reagents, e.g. for expression of cytokines,
CC receptors and growth factors, or to increase the copy number of a gene in
CC a cell. The constructs may also be used for functional and structural
CC analysis of chromosomes
XX

SQ Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

Query Match 100.0%; Score 2001; DB 2; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78595 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 78654

Qy         61 TCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78655 TCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAA 78714

Qy        121 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78715 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAAC 78774

Qy        181 AAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78775 AAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCT 78834

Qy        241 GTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78835 GTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGA 78894

Qy        301 TGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78895 TGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTTT 78954

Qy        361 TTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78955 TTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAAC 79014

Qy        421 ATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79015 ATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTA 79074

Qy        481 CTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79075 CTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAG 79134

Qy        541 CTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79135 CTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTT 79194

Qy        601 TTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79195 TTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTT 79254

Qy        661 TTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79255 TTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCA 79314

Qy        721 GGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79315 GGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAG 79374

Qy        781 CATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79375 CATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTG 79434

Qy        841 ACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79435 ACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACAC 79494

Qy        901 GAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79495 GAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGCT 79554

Qy        961 GTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79555 GTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCAC 79614
```

```
Qy       1021 TAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGAC 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79615 TAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGAC 79674

Qy       1081 CTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79675 CTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTC 79734

Qy       1141 CATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79735 CATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACA 79794

Qy       1201 GTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAAT 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79795 GTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAAT 79854

Qy       1261 GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79855 GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCA 79914

Qy       1321 TAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAAT 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79915 TAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAAT 79974

Qy       1381 ACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79975 ACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCTT 80034

Qy       1441 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80035 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 80094

Qy       1501 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80095 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 80154

Qy       1561 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80155 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 80214

Qy       1621 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80215 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 80274

Qy       1681 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80275 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 80334

Qy       1741 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80335 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 80394

Qy       1801 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTAC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80395 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTAC 80454

Qy       1861 AGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80455 AGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACT 80514

Qy       1921 TTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80515 TTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGT 80574

Qy       1981 GAAGTCATTCTGCCAGAATTC 2001
             |||||||||||||||||||||
Db      80575 GAAGTCATTCTGCCAGAATTC 80595
```

RESULT 2
AAV83942
ID AAV83942 standard; DNA; 18443 BP.
XX
AC AAV83942;
XX
DT 03-MAR-1999 (first entry)
XX
DE Bacterial artificial chromosome (BAC)-F2 contig 2.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; BAC; transgenic;
KW human artificial chromosome; bacterial artificial chromosome; ss.
XX
OS Synthetic.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 10; Page 181-194; 540pp; English.
XX
CC The present sequence represents a bacterial artificial chromosome (BAC)
CC contig, and exemplifies the invention. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in a
CC compatible cell, of replicating, acting as extra-chromosomal element and
CC segregating during cell division. The sequences can be used to construct
CC artificial chromosomes for use in gene therapy comprising a replicable,
CC segregating nucleic acid that confers a specific phenotype on cells.
CC Human artificial chromosomes can propagate in human cells and carry large
CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC they are not mutagenic. The artificial chromosomes are also useful for
CC generation of transgenic plants and animals, in production of proteins
CC and to make diagnostic reagents, e.g. for expression of cytokines,
CC receptors and growth factors, or to increase the copy number of a gene in
CC a cell. The constructs may also be used for functional and structural
CC analysis of chromosomes
XX
SQ Sequence 18443 BP; 5380 A; 3651 C; 3772 G; 5627 T; 0 U; 13 Other;

Query Match 99.2%; Score 1984.8; DB 2; Length 18443;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1997; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

```
Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       4292 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATTGT 4351

Qy         61 TCAGGA-CCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 119
              |||| | || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4352 TCAGAACCCTTGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 4411

Qy        120 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4412 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 4471

Qy        180 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4472 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 4531

Qy        240 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4532 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 4591

Qy        300 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4592 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 4651

Qy        360 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4652 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 4711

Qy        420 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4712 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 4771

Qy        480 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4772 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 4831

Qy        540 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4832 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 4891

Qy        600 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4892 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 4951

Qy        660 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4952 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 5011

Qy        720 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5012 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 5071

Qy        780 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5072 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 5131

Qy        840 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5132 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 5191

Qy        900 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5192 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 5251

Qy        960 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5252 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 5311

Qy       1020 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5312 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 5371

Qy       1080 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5372 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 5431

Qy       1140 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5432 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 5491

Qy       1200 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5492 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 5551

Qy       1260 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5552 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 5611

Qy       1320 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5612 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 5671

Qy       1380 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5672 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 5731

Qy       1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5732 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 5791

Qy       1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5792 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 5851

Qy       1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5852 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 5911

Qy       1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5912 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 5971

Qy       1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 1739
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5972 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 6031

Qy       1740 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 1799
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6032 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 6091

Qy       1800 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 1859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6092 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 6151

Qy       1860 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGAC 1919
              ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db       6152 CAGGTGCCAAGTTAAGTGCCCAAGCTTGMTCTTAAAAGCATACTGGATTTTGTTTTAGAC 6211

Qy       1920 TTTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGG 1979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6212 TTTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGG 6271

Qy       1980 TGAAGTCATTCTGCCAGAATTC 2001
              |||||||||||||||||||||
Db       6272 TGAAGTCATTCTGCCAGAATTC 6293
```

```
RESULT 3
AAV83940
ID   AAV83940 standard; DNA; 80240 BP.
XX
AC   AAV83940;
XX
DT   03-MAR-1999  (first entry)
XX
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
```

PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 9; Fig 16A; 540pp; English.
XX
CC The present sequence represents the NC-contig derived from a mutated
CC human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC mitotically stable and contains a functional neocentromere at a location
CC regarded as non-centromeric. This neocentromere maps to q25.2 on
CC chromosome 10. The specification describes nucleic acid sequences derived
CC from a eukaryotic chromosome, including a neocentromere or its functional
CC derivative or hybrid, that are able, in a compatible cell, of
CC replicating, acting as extra-chromosomal element and segregating during
CC cell division. The sequences can be used to construct artificial
CC chromosomes for use in gene therapy comprising a replicable, segregating
CC nucleic acid that confers a specific phenotype on cells. Human artificial
CC chromosomes can propagate in human cells and carry large amounts of DNA
CC (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC mutagenic. The artificial chromosomes are also useful for generation of
CC transgenic plants and animals, in production of proteins and to make
CC diagnostic reagents, e.g. for expression of cytokines, receptors and
CC growth factors, or to increase the copy number of a gene in a cell. The
CC constructs may also be used for functional and structural analysis of
CC chromosomes
XX
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

Query Match 94.7%; Score 1895.2; DB 2; Length 80240;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1908; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      78329 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATTGT 78388

Qy         61 TCAGGA-CCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 119
              |||| | || ||||||||||||||||||||||||||||||||||||||||||||||||
Db      78389 TCAGAACCCTTGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 78448

Qy        120 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78449 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 78508

Qy        180 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78509 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 78568

Qy        240 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78569 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 78628

Qy        300 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78629 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 78688

Qy        360 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78689 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 78748

Qy        420 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 479
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      78749 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 78808

Qy        480 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78809 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 78868

Qy        540 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78869 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 78928

Qy        600 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78929 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 78988

Qy        660 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78989 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 79048

Qy        720 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79049 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 79108

Qy        780 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79109 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 79168

Qy        840 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79169 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 79228

Qy        900 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79229 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 79288

Qy        960 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79289 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 79348

Qy       1020 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79349 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 79408

Qy       1080 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79409 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 79468

Qy       1140 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79469 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 79528

Qy       1200 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79529 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 79588

Qy       1260 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79589 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 79648

Qy       1320 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79649 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 79708

Qy       1380 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79709 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 79768

Qy       1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79769 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 79828

Qy       1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79829 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 79888
```

```
Qy       1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79889 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 79948

Qy       1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79949 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 80008

Qy       1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 1739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80009 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 80068

Qy       1740 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 1799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80069 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 80128

Qy       1800 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 1859
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80129 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 80188

Qy       1860 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 1911
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80189 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 80240
```

RESULT 4
AAV83958
ID AAV83958 standard; DNA; 300 BP.
XX
AC AAV83958;
XX
DT 03-MAR-1999 (first entry)
XX
DE Bacterial artificial chromosome (BAC)-F2 contig 47 fragment 2.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; BAC; transgenic;
KW human artificial chromosome; bacterial artificial chromosome; ss.
XX
OS Synthetic.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 10; Page 235; 540pp; English.
XX
CC The present sequence represents a bacterial artificial chromosome (BAC)
CC contig, and exemplifies the invention. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in a
CC compatible cell, of replicating, acting as extra-chromosomal element and
CC segregating during cell division. The sequences can be used to construct
CC artificial chromosomes for use in gene therapy comprising a replicable,
CC segregating nucleic acid that confers a specific phenotype on cells.
CC Human artificial chromosomes can propagate in human cells and carry large
CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC they are not mutagenic. The artificial chromosomes are also useful for
CC generation of transgenic plants and animals, in production of proteins
CC and to make diagnostic reagents, e.g. for expression of cytokines,
CC receptors and growth factors, or to increase the copy number of a gene in
CC a cell. The constructs may also be used for functional and structural
CC analysis of chromosomes
XX
SQ Sequence 300 BP; 54 A; 70 C; 56 G; 101 T; 0 U; 19 Other;

```
  Query Match             4.9%;  Score 97.6;  DB 2;  Length 300;
  Best Local Similarity   66.7%;  Pred. No. 3.4e-15;
  Matches  168;  Conservative   18;  Mismatches   62;  Indels    4;  Gaps    4;

Qy       1753 GAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTG 1812
             ||||||| | ::|| |:|||||||||:||||| | :||   |           |||||||
Db          1 GAATTCTCTGWKATTAKAACTATCTTGMCTCAAATTSACTTGGTGAGCTAACCTGGCCTG 60

Qy       1813 TGGTCCCTACG-TAGAAAGGAGGCTTTGTCATAAAG-TCTTATATGGTACAGGTGCCAAG 1870
             ||||||||  | |  || |||||||||||||||| || || |:| |||||| :|||| ||
Db         61 TGGTCCCTTGGCTTTAATGGAGGCTTTGTCATATAGATCATMTGTGGTACTKGTGCCTAG 120

Qy       1871 TT-AAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTTAGTGAA 1929
             ||  |||||||   ||||||:||  : || ||||:|||||:|   || || | :| : ||
Db        121 TTGTAGTGCCCTGCCTTGCTSTTCTWGGCTTACTKGATTTWGGGGTATACATCWATKTAA 180

Qy       1930 CTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGTGAAGTCATT 1989
             :|:|| ||  |   |   |||| ::||||||| |||||||||  ||| || ||| || ||
Db        181 YTSAAAGGTCT-TTCTCCTCCCGYYGGGAGAATTTCTCCTCCTCCCTCGGAGAACTCTTT 239

Qy       1990 CTGCCAGAATTC 2001
             ||:||  |||||
Db        240 CTSCCGAAATTC 251
```

RESULT 5
ADA71938
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX

```
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1
CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,
CC   or in a corresponding incompatible or compatible interaction. (M1) is
CC   useful for conferring resistance to resistance or tolerance to a plant to
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

  Query Match             2.9%;  Score 58.8;  DB 8;  Length 2000;
  Best Local Similarity   11.8%;  Pred. No. 0.00013;
  Matches   54;  Conservative  215;  Mismatches  187;  Indels    2;  Gaps    1;

Qy       1361 ATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTA 1420
Db        508 MYRMRWKRYYMSARYTMRYCARKKYSYSAARKARCWYRGKGYYWAGMWMKRYKRMYMYKM 567

Qy       1421 TTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTAC 1480
Db        568 MWWYKRKYSKCSWYCKMSYYASCMKSARKAGAKMCKRSKMSAWSKSMRSSRKCRKCASKR 627

Qy       1481 ATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGT 1540
Db        628 SSAKRYAMMGGMTSGSRMSRWKSYTCYWRKWGSMKSTCTWMYYMSKYTYAKYGSYWRYRY 687

Qy       1541 AAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCA 1600
Db        688 RAWCMYMWRWYYYRYRSYMTYMAWYTSSTRMAMTGMKYSGRYWTSWYKYCKCSWKYRSMW 747

Qy       1601 TAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTG 1660
Db        748 YYWSWWWAKTWKMWRRYATRMMWMWYRYSMKWYTWCTMWGYWWYWWRTYMKMRYMW--YK 805

Qy       1661 GTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAA 1720
Db        806 CTKTYWYWSATYWTGTWAAWWMAKTKMRMGMTGAKTRGRARKARYWWKWATWCATKRWMT 865

Qy       1721 TCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTCA 1780
Db        866 KGKGAKWAWTWMAKAWRKYYWSWMRAWYYYYKTRRTRYKTCWWKARWGSWAYWRMWWKGS 925

Qy       1781 CTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGTCC 1818
Db        926 AKMWWMWKGGRWGWTKYWYWYCTTWKMACGRATKYMCC 963


RESULT 6
ACL37108
ID   ACL37108 standard; cDNA; 2000 BP.
XX
AC   ACL37108;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:15671.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 15671; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 510 A; 350 C; 268 G; 518 T; 0 U; 354 Other;

  Query Match             2.6%;  Score 52;  DB 11;  Length 2000;
  Best Local Similarity   12.5%;  Pred. No. 0.0079;
  Matches   49;  Conservative  174;  Mismatches  169;  Indels    0;  Gaps    0;

Qy       1356 TTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACAT 1415
Db         39 TKYYKWYYYRTCSYKRWWMTSMMKYRAYMMRWYWKKMWAMWYAWCMWYAYGYMMWKWWYS 98

Qy       1416 ATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGG 1475
Db         99 RTMTYTRAYYKKWKMWRRKMSRMAAAWYYWSYWRTSRWKWTGYRKKRWWRARRYMYWTYW 158

Qy       1476 CTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCA 1535
Db        159 CYTWSWMRWYRYWKMMKYRAYKWMAYTTTYMTKCCMWWWCTYGYKWKWAAWRASAKRARR 218

Qy       1536 TATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACT 1595
Db        219 AWMMMRWKSMWTWAATWSWWSMMATWKRKMRWGWMWAYSSTWMMRWWWYCCTRRAMYKMK 278

Qy       1596 TATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACA 1655
Db        279 AATMTRCTWYMWMAAMTSAWRAWCAGWATSYATKWYAWRTTYMKYRRRYYACMAWTRWTK 338

Qy       1656 AGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTG 1715
Db        339 MWSMMRTKKWWYTMARSWGCKKWMWAKARMATMCMAYTYSTYRYYAWRRRWMRWSWAAWY 398

Qy       1716 TTTAATCAGGTCATTGAGAATCTATTAAAAAG 1747
Db        399 SAKYWYAKSYYYMYKCARKWWTTYAYYAAAAK 430


RESULT 7
AAC15209
ID   AAC15209 standard; cDNA; 431 BP.
XX
AC   AAC15209;
XX
DT   06-OCT-2000  (first entry)
```

```
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 19284.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 19284; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 431 BP; 109 A; 77 C; 112 G; 132 T; 0 U; 1 Other;

  Query Match             2.5%;  Score 49.4;  DB 3;  Length 431;
  Best Local Similarity   71.4%;  Pred. No. 0.02;
  Matches   65;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         70 TGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATATTG 129
              | | ||| |||||||||||||| ||| |||||  | | || | ||||| || ||||||| |
Db        231 TTGGTTGTAAGTGACAGAAACCCAACTCAAATTAGCTTAGGCCAAAAAGAGAATATATGG 290

Qy        130 GCTCATGTAACCTTCTCACAGAGAGGGCAGG 160
              | ||||| ||||    |  ||| | || | |
Db        291 GTTCATGGAACCCAAACTAAGAAAAGGGACG 321


RESULT 8
ADZ70899/c
ID   ADZ70899 standard; DNA; 1428 BP.
XX
AC   ADZ70899;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Human matrix attachment region DNA SEQ ID NO 23.
XX
KW   ds; matrix attachment region; MAR; protein production.
XX
OS   Homo sapiens.
XX
PN   WO2005040377-A2.
XX
PD   06-MAY-2005.
XX
PF   22-OCT-2004; 2004WO-EP011974.
XX
PR   24-OCT-2003; 2003US-0513574P.
PR   06-FEB-2004; 2004EP-00002722.
XX
PA   (SELE-) SELEXIS SA.
XX
PI   Mermod N,  Girod PA,  Bucher P,  Nguyen D,  Calabrese D,  Saugy D;
PI   Puttini S;
XX
DR   WPI; 2005-333507/34.
XX
PT   New purified and isolated DNA sequence having protein production
PT   increasing activity comprises a bent DNA element and a binding site for a
PT   DNA binding protein, useful for increasing protein production activity in
PT   eukaryotic host cell.
XX
PS   Claim 4; SEQ ID NO 23; 282pp; English.
XX
CC   The invention relates to a purified and isolated DNA sequence having
CC   protein production increasing activity comprising at least one bent DNA
CC   element, and at least one binding site for a DNA binding protein. The
CC   purified and isolated DNA sequence comprising a first and a second
CC   isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC   MAR nucleotide sequence selected from a purified and isolated DNA
CC   sequence above, a purified and isolated MAR DNA above, a purified and
CC   isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC   sequence complementary to it, its molecular chimera, or its combinations
CC   and variants, is useful for increasing protein production activity in a
CC   eukaryotic host cell. The present sequence represents a human matrix
CC   attachment region DNA.
XX
SQ   Sequence 1428 BP; 652 A; 18 C; 6 G; 752 T; 0 U; 0 Other;

  Query Match             2.5%;  Score 49.4;  DB 14;  Length 1428;
  Best Local Similarity   46.0%;  Pred. No. 0.033;
  Matches  210;  Conservative    0;  Mismatches  241;  Indels    6;  Gaps    1;

Qy       1310 ATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCAT 1369
              | | || |  ||| | ||  | |||  | | |||| ||  |  | | |  |||  | |
Db       1395 AATATAATAAATATTAATAATTAATAAATTAATAAATTTAATTAATATATTATATATAAA 1336

Qy       1370 CTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAG 1429
              || | |  |||   | | ||  |  |  |   | |  ||||| ||      ||  | |
Db       1335 ATATATATTATAATATATTATATATAAAATATATATTATAATATATTATATATAAAATAT 1276

Qy       1430 CTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAG 1489
              | |  |  |  |   || ||    |||||    |||       |   ||| ||||||||
Db       1275 ATATTATAATATATTATATATAAAATATATATTATAA------TATATTATATATAAAAT 1222

Qy       1490 GGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTC 1549
              ||||     || | ||| | ||    | | |     |  | | |||| ||||  ||
Db       1221 ATATATTATAATATATTATATATAAAATATATATTATAATATATTATATATAAAATATAT 1162

Qy       1550 ATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGT 1609
              |||| | |   | | ||| |           || |      | |  |||        ||
Db       1161 ATTATAATATATTATATATAAAATATATATTATATTATATATAAAATATATATTATAATA 1102

Qy       1610 CATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTC 1669
              ||||  ||   || | | | || ||| |||   ||   ||  | |   |    |   |
Db       1101 TATTATATATAAAATATATATTATAATATATTATATATAAAATATATATTATAATATATT 1042

Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
              |     ||||||   |  | ||   || |    |     | ||  |   | ||  |  | |
Db       1041 ATATATAAAATATATATTATAATATATATTATAATATATTATATATAAAATATATATT 982

Qy       1730 TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGAT 1766
              |  || |||   | | ||| | |  | |  |  ||
Db        981 ATAATATATATATTATAATATATTATATATAAAATAT 945
```

```
RESULT 9
ADZ70899
ID   ADZ70899 standard; DNA; 1428 BP.
XX
AC   ADZ70899;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Human matrix attachment region DNA SEQ ID NO 23.
XX
KW   ds; matrix attachment region; MAR; protein production.
XX
OS   Homo sapiens.
XX
PN   WO2005040377-A2.
XX
PD   06-MAY-2005.
XX
PF   22-OCT-2004; 2004WO-EP011974.
XX
PR   24-OCT-2003; 2003US-0513574P.
PR   06-FEB-2004; 2004EP-00002722.
XX
PA   (SELE-) SELEXIS SA.
XX
PI   Mermod N,  Girod PA,  Bucher P,  Nguyen D,  Calabrese D,  Saugy D;
PI   Puttini S;
XX
DR   WPI; 2005-333507/34.
XX
PT   New purified and isolated DNA sequence having protein production
PT   increasing activity comprises a bent DNA element and a binding site for a
PT   DNA binding protein, useful for increasing protein production activity in
PT   eukaryotic host cell.
XX
PS   Claim 4; SEQ ID NO 23; 282pp; English.
XX
CC   The invention relates to a purified and isolated DNA sequence having
CC   protein production increasing activity comprising at least one bent DNA
CC   element, and at least one binding site for a DNA binding protein. The
CC   purified and isolated DNA sequence comprising a first and a second
CC   isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC   MAR nucleotide sequence selected from a purified and isolated DNA
CC   sequence above, a purified and isolated MAR DNA above, a purified and
CC   isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC   sequence complementary to it, its molecular chimera, or its combinations
CC   and variants, is useful for increasing protein production activity in a
CC   eukaryotic host cell. The present sequence represents a human matrix
CC   attachment region DNA.
XX
SQ   Sequence 1428 BP; 652 A; 18 C; 6 G; 752 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 45;  DB 14;  Length 1428;
  Best Local Similarity   43.5%;  Pred. No. 0.48;
  Matches  204;  Conservative    0;  Mismatches  265;  Indels    0;  Gaps    0;

Qy       1310 ATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCAT 1369
Db        786 ATTATAATATATTATAATATATATTTTATATATAATATATTATAATATATATTTTATATA 845

Qy       1370 CTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAG 1429
Db        846 TAATATATTATAATATATATTTTATATATAATATATTATAATATATATATTATAATATAT 905

Qy       1430 CTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAG 1489
Db        906 ATTTTATATATAATATATTATCATATATATATTAAATATATATTTTATATATAATATATT 965

Qy       1490 GGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTC 1549
Db        966 ATAATATATATATTATAATATATATTTTATATATAATATATTATAATATATATATTATAA 1025

Qy       1550 ATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGT 1609
Db       1026 TATATATTTTATATATAATATATTATAATATATATTTTATATATAATATATTATAATATA 1085

Qy       1610 CATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTC 1669
Db       1086 TATTTTATATATAATATATTATAATATATATTTTATATATAATATAATATATATTTTATA 1145

Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
Db       1146 TATAATATATTATAATATATATTTTATATATAATATATTATAATATATATTTTATATATA 1205

Qy       1730 TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTT 1778
Db       1206 ATATATTATAATATATATTTTATATATAATATATTATAATATATATTTT 1254


RESULT 10
ABK28365
ID   ABK28365 standard; DNA; 9238 BP.
XX
AC   ABK28365;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA transcription associated genomic DNA #120.
XX
KW   DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW   PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW   single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW   viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW   immunological disorder; Werner syndrome; developmental disorder;
KW   psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW   neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW   myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW   angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW   polyglutamine disorder; solid tumour.
XX
OS   Unidentified.
XX
PN   WO200192565-A2.
XX
PD   06-DEC-2001.
XX
PF   06-APR-2001; 2001WO-EP003973.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-090046/12.
XX
PT   New nucleic acids or oligomers, useful for diagnosing or treating
PT   diseases associated with DNA transcription, e.g. immunological disorders,
PT   Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 239; 32pp; English.
XX
CC   The invention relates to a nucleic acid, which comprises a segment of the
CC   chemically pretreated DNA of genes associated with DNA transcription from
CC   one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC   or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC   to the chemically pretreated DNA of genes associated with DNA
CC   transcription. The set of oligomer probes are useful for detecting the
CC   cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC   in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC   diagnosing or treating diseases associated with DNA transcription
```

```
CC   (particularly with the methylation status), e.g. adenosine deaminase
CC   deficiency, viral infection, retroviral infection, Sezary syndrome,
CC   haematological disorders, immunological disorders, Werner syndrome,
CC   tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC   neurological disorders, neurodegenerative disorders, Waardenburg
CC   syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC   infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC   disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC   or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC   associated genomic DNA molecules of the invention. Note: The sequence
CC   data for this patent did not form part of the printed specification but
CC   was obtained in electronic format directly from the European Patent
CC   Office
XX
SQ   Sequence 9238 BP; 2692 A; 159 C; 1985 G; 4402 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 44.8;  DB 6;  Length 9238;
  Best Local Similarity   56.9%;  Pred. No. 1.2;
  Matches   82;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy       1611 ATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCA 1670
             || |  ||   ||  |||| ||  ||  || ||   | |    ||  || ||| | |
Db       1932 ATAATTTATAAAAAGAAAAATAATATTGATATAGGAAATATGTAAAAATGTTTTATGTTA 1991

Qy       1671 ATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATT 1730
              |  ||||||||||||| | |||  |     || |  |||||| | ||| || |  | |||
Db       1992 TTTAAAAAATAAGAAATATAAATTAAAATTAAATTTAAATATATTTTTTATTATTTGATT 2051

Qy       1731 GAGAATCTATTAAAAAGTATTTGA 1754
              |  ||  | |||||||   | | |
Db       2052 GGTAAAATTTTAAAAATGTTGTAA 2075


RESULT 11
AAV20419/c
ID   AAV20419 standard; cDNA to mRNA; 3100 BP.
XX
AC   AAV20419;
XX
DT   15-JUN-1998  (first entry)
XX
DE   Human discs large 1 gene-1 cDNA cancer related molecule.
XX
KW   Human; NE-dlg; discs large 1 gene; cancer related molecule; nerve;
KW   internal secretion tissue; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             234. .2687
FT                   /*tag=  a
FT                   /product= "dlg 1"
XX
PN   JP10066581-A.
XX
PD   10-MAR-1998.
XX
PF   23-AUG-1996;   96JP-00241370.
XX
PR   23-AUG-1996;   96JP-00241370.
XX
PA   (SUME ) SUMITOMO ELECTRIC IND CO.
XX
DR   WPI; 1998-224339/20.
DR   P-PSDB; AAW48101.
XX
PT   Human discs large 1 gene family - useful in, e.g. therapeutic
PT   composition(s) for treating cancer.
XX
PS   Claim 16; Page 17-18; 31pp; Japanese.
XX
CC   The present sequence encodes human dlg-1 molecule. The present invention
CC   describes human discs large 1 gene (dlg) family expressible in nerve
CC   tissue. Also described are: (1) a polynucleotide (PN) encoding dlg and
CC   comprising a 3100 (I) or 3213 bp (II) sequence; (2) RNA corresponding to
CC   PN having sequences (I) and (II); (3) RNA having a base sequence
CC   translated by dlg; (4) an antisense PN having a at least 15 bp sequence,
CC   and which is a part of PN of (1); (5) a derivative of the antisense PN;
CC   (6) an antibody specific for dlg, and (7) an antibody specific for a
CC   polypeptide having a 817 or 849 aa sequence. The polypeptide, RNA's and
CC   antibodies can be used for detection of dlg. The antisense PN can be used
CC   as a therapeutic composition for treating cancer
XX
SQ   Sequence 3100 BP; 753 A; 799 C; 902 G; 646 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 44.4;  DB 2;  Length 3100;
  Best Local Similarity   48.4%;  Pred. No. 0.96;
  Matches  123;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
              ||  |||||| || ||| |  |     |   ||| | ||  | |||    |       |
Db       2859 AACTTAAAAAAAAAAAAAGAACAGGACTAGAAAACTTAATTTTTTTTGACATATCTAAAAA 2800

Qy       1730 TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGG 1789
              | | |          |  || |||||    ||   |||  || |   ||   ||   | | ||
Db       2799 AGGGGACAAGAACAGTAGCATTTGTTCTCCCCATGGGGACAGGAATGAAGAGGGAGGAGG 2740

Qy       1790 ACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTC 1849
                || ||      |||   ||||   |   |     |   |||||   |      | |||
Db       2739 GAAGAGGGACTTGATTTCTTCTGTTCACAAGAGAATGGTTGGAGGGGATTCTTCAGAGTT 2680

Qy       1850 TTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTT 1909
              ||  | || |  ||  |||| |  |||||||  || |  || ||      |  | |||||
Db       2679 TTTCAGGGGATGGGACCCAAATGTAGTGCCCAGACTGGTCCTCAATGATTTGTTTGATTT 2620

Qy       1910 TGTTTTAGACTTTT 1923
              |||| || |  | |
Db       2619 TGTTATAAATCTCT 2606


RESULT 12
AAX86725/c
ID   AAX86725 standard; DNA; 3100 BP.
XX
AC   AAX86725;
XX
DT   27-OCT-1999  (first entry)
XX
DE   DNA sequence encoding tumour suppressor protein NE-dlg.
XX
KW   Nedasin; tumour suppressor protein; NE-dlg; tumour suppression;
KW   malignant tumour formation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             234. .2687
FT                   /*tag=  a
XX
PN   WO9943702-A1.
XX
PD   02-SEP-1999.
XX
PF   24-AUG-1998;   98WO-JP003740.
XX
PR   25-FEB-1998;   98JP-00043552.
XX
PA   (SUME ) SUMITOMO ELECTRIC IND CO.
XX
PI   Kishimoto T,  Niwa S,  Kuwahara H,  Saya H;
XX
DR   WPI; 1999-518835/43.
DR   P-PSDB; AAY30137.
XX
PT   New nedasin protein, useful in the investigation of the mechanism of
PT   formation of malignant tumors.
```

```
XX
PS   Example 1; Page 87-96; 104pp; Japanese.
XX
CC   The present sequence encodes tumour suppressor protein NE-dlg. The
CC   specification describes mammalian proteins, designated nedasin, which
CC   bind to the tumour suppressor protein NE-dlg. Nedasin is involved with
CC   the regulation of tumour suppression mechanisms. Nedasin and antibodies
CC   recognizing it are useful in the investigation of the mechanism of
CC   formation of malignant tumours. The polynucleotides are useful for
CC   screening cDNA libraries, and as a source of primers and probes
XX
SQ   Sequence 3100 BP; 753 A; 799 C; 902 G; 646 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 44.4;  DB 2;  Length 3100;
  Best Local Similarity   48.4%;  Pred. No. 0.96;
  Matches  123;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
              ||  |||||| || ||| |  |     |   ||| | ||  | |||     |        |
Db       2859 AACTTAAAAAAAAAAAAAGAACAGGACTAGAAAACTTAATTTTTTTTGACATATCTAAAAA 2800

Qy       1730 TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGG 1789
              | | |         |  || |||||   ||   |||  || |   ||   ||   | | ||
Db       2799 AGGGGACAAGAACAGTAGCATTTGTTCTCCCCATGGGGACAGGAATGAAGAGGGAGGAGG 2740

Qy       1790 ACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTC 1849
                || ||       |||   ||||  |   |     |   |||||   |     | |||
Db       2739 GAAGAGGGACTTGATTTCTTCTGTTCACAAGAGAATGGTTGGAGGGGATTCTTCAGAGTT 2680

Qy       1850 TTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTT 1909
              ||  | ||  |  ||  |||| |  |||||||  || |  || ||      |  | |||||
Db       2679 TTTCAGGGGATGGGACCCAAATGTAGTGCCCAGACTGGTCCTCAATGATTTGTTTGATTT 2620

Qy       1910 TGTTTTAGACTTTT 1923
              |||| || |  | |
Db       2619 TGTTATAAATCTCT 2606


RESULT 13
ACC85073/c
ID   ACC85073 standard; DNA; 3100 BP.
XX
AC   ACC85073;
XX
DT   13-OCT-2003  (first entry)
XX
DE   Human MBCAT polypeptide encoding DNA.
XX
KW   MBCAT; beta-catenin; cytostatic; gene therapy; cancer; human; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             234. .2687
FT                   /*tag=  a
FT                   /product= "MBCAT"
XX
PN   WO2003052068-A2.
XX
PD   26-JUN-2003.
XX
PF   12-DEC-2002; 2002WO-US039796.
XX
PR   13-DEC-2001; 2001US-0340213P.
PR   13-DEC-2001; 2001US-0340314P.
PR   13-DEC-2001; 2001US-0340322P.
PR   15-FEB-2002; 2002US-0357502P.
XX
PA   (EXEL-) EXELIXIS INC.
XX
PI   Costa MA,  Gendreau SB,  Dora EG,  Nicoll M;
XX
DR   WPI; 2003-533010/50.
DR   P-PSDB; ABR82222.
XX
PT   Identifying a candidate beta-catenin pathway modulating agent for
PT   diagnosing or treating cancer by detecting a test agent-biased activity
PT   of the assay system comprising a purified MBCAT polypeptide or nucleic
PT   acid.
XX
PS   Example; Page 53-54; 81pp; English.
XX
CC   The invention relates to genes that modify beta-catenin pathway and to
CC   the identification of human MBCAT (modifiers of beta-catenin) genes. The
CC   MBCAT polypeptides are therapeutic targets for disorders associated with
CC   beta-catenin function and are useful for manufacturing a medicament for
CC   diagnosing or treating breast, colon, lung or ovary cancer. The present
CC   sequence represents a human MBCAT polypeptide encoding DNA
XX
SQ   Sequence 3100 BP; 753 A; 799 C; 902 G; 646 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 44.4;  DB 9;  Length 3100;
  Best Local Similarity   48.4%;  Pred. No. 0.96;
  Matches  123;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
              ||  |||||| || ||| |  |     |   ||| | ||  | |||     |        |
Db       2859 AACTTAAAAAAAAAAAAAGAACAGGACTAGAAAACTTAATTTTTTTTGACATATCTAAAAA 2800

Qy       1730 TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGG 1789
              | | |         |  || |||||   ||   |||  || |   ||   ||   | | ||
Db       2799 AGGGGACAAGAACAGTAGCATTTGTTCTCCCCATGGGGACAGGAATGAAGAGGGAGGAGG 2740

Qy       1790 ACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTC 1849
                || ||       |||   ||||  |   |     |   |||||   |     | |||
Db       2739 GAAGAGGGACTTGATTTCTTCTGTTCACAAGAGAATGGTTGGAGGGGATTCTTCAGAGTT 2680

Qy       1850 TTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTT 1909
              ||  | ||  |  ||  |||| |  |||||||  || |  || ||      |  | |||||
Db       2679 TTTCAGGGGATGGGACCCAAATGTAGTGCCCAGACTGGTCCTCAATGATTTGTTTGATTT 2620

Qy       1910 TGTTTTAGACTTTT 1923
              |||| || |  | |
Db       2619 TGTTATAAATCTCT 2606


RESULT 14
AAV20420/c
ID   AAV20420 standard; cDNA to mRNA; 3196 BP.
XX
AC   AAV20420;
XX
DT   15-JUN-1998  (first entry)
XX
DE   Human discs large 1 gene-2 cDNA cancer related molecule.
XX
KW   Human; NE-dlg; discs large 1 gene; cancer related molecule; nerve;
KW   internal secretion tissue; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             234. .2803
FT                   /*tag=  a
FT                   /product= "dlg 2"
XX
PN   JP10066581-A.
XX
PD   10-MAR-1998.
XX
PF   23-AUG-1996;   96JP-00241370.
XX
PR   23-AUG-1996;   96JP-00241370.
XX
PA   (SUME ) SUMITOMO ELECTRIC IND CO.
XX
```

DR WPI; 1998-224339/20.
DR P-PSDB; AAW48102.
XX
PT Human discs large 1 gene family - useful in, e.g. therapeutic
PT composition(s) for treating cancer.
XX
PS Claim 17; Page 20-21; 31pp; Japanese.
XX
CC The present sequence encodes human dlg-2 molecule. The present invention
CC describes human discs large 1 gene (dlg) family expressible in nerve
CC tissue. Also described are: (1) a polynucleotide (PN) encoding dlg and
CC comprising a 3100 (I) or 3213 bp (II) sequence; (2) RNA corresponding to
CC PN having sequences (I) and (II); (3) RNA having a base sequence
CC translated by dlg; (4) an antisense PN having a at least 15 bp sequence,
CC and which is a part of PN of (1); (5) a derivative of the antisense PN;
CC (6) an antibody specific for dlg, and (7) an antibody specific for a
CC polypeptide having a 817 or 849 aa sequence. The polypeptide, RNA's and
CC antibodies can be used for detection of dlg. The antisense PN can be used
CC as a therapeutic composition for treating cancer
XX
SQ Sequence 3196 BP; 793 A; 825 C; 920 G; 658 T; 0 U; 0 Other;

Query Match 2.2%; Score 44.4; DB 2; Length 3196;
Best Local Similarity 48.4%; Pred. No. 0.97;
Matches 123; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
             ||  |||||| || ||| |  |     |   ||| | ||  | |||    |       |
Db       2955 AACTTAAAAAAAAAAAAGAACAGGACTAGAAAACTTAATTTTTTTTGACATATCTAAAAA 2896

Qy       1730 TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGG 1789
              | | |     |  || |||||   ||   |||  || |   ||   ||   | | ||
Db       2895 AGGGGACAAGAACAGTAGCATTTGTTCTCCCCATGGGGACAGGAATGAAGAGGGAGGAGG 2836

Qy       1790 ACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTC 1849
               || ||       |||   ||||   |   |     |   |||||   |     | |||
Db       2835 GAAGAGGGACTTGATTTCTTCTGTTCACAAGAGAATGGTTGGAGGGGATTCTTCAGAGTT 2776

Qy       1850 TTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTT 1909
             ||  | || |  ||  |||| |  |||||||  || |  || ||     |  | |||||
Db       2775 TTTCAGGGGATGGGACCCAAATGTAGTGCCCAGACTGGTCCTCAATGATTTGTTTGATTT 2716

Qy       1910 TGTTTTAGACTTTT 1923
              |||| || |  | |
Db       2715 TGTTATAAATCTCT 2702
```

RESULT 15
ABV25006/c
ID ABV25006 standard; cDNA; 5147 BP.
XX
AC ABV25006;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 24997.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 4841-4843; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ Sequence 5147 BP; 1261 A; 1238 C; 1454 G; 1188 T; 0 U; 6 Other;

Query Match 2.2%; Score 44.4; DB 5; Length 5147;
Best Local Similarity 48.4%; Pred. No. 1.2;
Matches 123; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
Qy       1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
             ||  |||||| || ||| |  |     |   ||| | ||  | |||    |       |
Db       2859 AACTTAAAAAAAAAAAAAGAACAGGACTAGAAAACTTAATTTTTTTTGACATATCTAAAAA 2800

Qy       1730 TGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGG 1789
              | | |       |  || |||||   ||   |||  || |   ||   ||   | | ||
Db       2799 AGGGGACAAGAACAGTAGCATTTGTTCTCCCCATGGGGACAGGAATGAAGAGGGAGGAGG 2740

Qy       1790 ACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTC 1849
               || ||       |||   ||||   |   |     |   |||||   |     | |||
Db       2739 GAAGAGGGACTTGATTTCTTCTGTTCACAAGAGAATGGTTGGAGGGGATTCTTCAGAGTT 2680

Qy       1850 TTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTT 1909
             ||  | || |  ||  |||| |  ||||||||  || |  || ||     |  | |||||
Db       2679 TTTCAGGGGATGGGACCCAAATGTAGTGCCCAGACTGGTCCTCAATGATTTGTTTGATTT 2620

Qy       1910 TGTTTTAGACTTTT 1923
              |||| || |  | |
Db       2619 TGTTATAAATCTCT 2606
```

RESULT 16
ADA73182
ID ADA73182 standard; DNA; 2000 BP.
XX
AC ADA73182;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 6508.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX

```
PF   22-JUN-2001; 2001WO-IB001105.
XX
PR   22-JUN-2001; 2001WO-IB001105.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Chang H,  Chen W,  Cooper B,  Glazebrook J,  Goff SA,  Hou Y;
PI   Katagiri F,  Quan S,  Tao Y,  Whitham S,  Xie Z,  Zhu T,  Zou G;
XX
DR   WPI; 2003-175290/17.
XX
PT   Identifying at least one gene involved in plant resistance or response to
PT   pathogenic infection for conferring resistance or tolerance to a plant to
PT   bacterial, fungal or viral infection by determining or detecting plant
PT   gene expression.
XX
PS   Claim 27; SEQ ID NO 6508; 899pp; English.
XX
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1
CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,
CC   or in a corresponding incompatible or compatible interaction. (M1) is
CC   useful for conferring resistance to resistance or tolerance to a plant to
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 2000 BP; 715 A; 429 C; 393 G; 463 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 44.2;  DB 8;  Length 2000;
  Best Local Similarity   56.6%;  Pred. No. 0.9;
  Matches   82;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy       1608 GTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGC 1667
              || | ||  |||    |||||   | |||  ||   |||  |||||  |  || |  ||
Db        167 GTAAATATATAAAATTCAAAATTAAAAATAAATAATATTGTAAGAAGAGTCTAGAGTTCA 226

Qy       1668 TCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTC 1727
              |  |||||||||  || |||  |  |  || |||||   ||||  |     ||||
Db        227 TGTATATAAAAAAAATAAATGGCTAAAATTTGAAACAAAAAAGAAAATATTAAAAGATGG 286

Qy       1728 ATTGAGAATCTATTAAAAAGTATTT 1752
              || ||| ||  | || || ||| |
Db        287 GTTTAGAGTCCTTAAAGAAATATAT 311


RESULT 17
ACL35363
ID   ACL35363 standard; cDNA; 2000 BP.
XX
AC   ACL35363;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:13926.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 13926; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 540 A; 344 C; 416 G; 428 T; 0 U; 272 Other;

  Query Match             2.2%;  Score 44.2;  DB 11;  Length 2000;
  Best Local Similarity   13.0%;  Pred. No. 0.9;
  Matches   37;  Conservative  130;  Mismatches  118;  Indels    0;  Gaps    0;

Qy       1236 GTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTT 1295
              | ::  :::|::   | : |:  :   ::| :||: : :: :  :  :::::  : : ::
Db         36 GAKWWTMWSAYMAMYAMWYAKKKWRMTRMAKRCTSCSKYMSWASKMRWWWWSMARGWSKW 95

Qy       1296 TCACAAGCCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAATT 1355
              |::  :   :  :::: :  :   |: :| ::::::| |:|     |: ::   | ::::
Db         96 TYMGGRMWRYKAYMWKRWMYMMAWAWSWTRMYWWWKACTYCYTMYKTMYWYRWAAAWRWW 155

Qy       1356 TTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCTTTGCTAATACAT 1415
              :::  ::|:|::|:::|:::: : | :  :   :  :      | : : :    ||::|:
Db        156 KWYKKWWARGWSAKMKAYRYRKWCTYYWAWRMAWRWMTTTKWCTWSWYAWWTCYATWMAW 215

Qy       1416 ATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGG 1475
              || : |: |:: :  |:  :|:  :|: : :::|::: ::   :  : ::  : :: ::
Db        216 ATAWTTWTTRMAWWKTRAWMTWWRYAWMYTWWWAKKWAWKYATKWTTYAWRWWMMMWMSR 275

Qy       1476 CTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTT 1520
               |:: :|: :: :  : : |: | :: :   :|   :::::: ::
Db        276 TTYRWMTWAMWYWTWSWYYTWAAMSWYWWCMWTTWSRRRKWYRYK 320


RESULT 18
ABV72623/c
ID   ABV72623 standard; DNA; 82938 BP.
XX
AC   ABV72623;
XX
DT   27-NOV-2002  (first entry)
XX
DE   Human transporter protein encoding gene.
XX
KW   Human; transporter protein; hTP; gamma-aminobutyric-acid; GABA;
KW   gene therapy; protein therapy; vaccine; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1120. .81932
```

```
FT                   /*tag= a
FT                   /product= "Transporter protein"
FT                   /note= "Contains 8 introns"
FT   exon            1120. .1223
FT                   /*tag= b
FT                   /number= 1
FT   intron          1224. .27733
FT                   /*tag= c
FT                   /number= 1
FT   variation       replace(2419,C)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3191,A)
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4195,T)
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5910,C)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7119,T)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7529,G)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8855,C)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8956,A)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10220,G)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11628,C)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11670,A)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11671,G)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11931,C)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15984,G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17735,A)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17910,A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18130,C)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19293. .19294,TWA)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19294. .19295,ATT)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19296. .19297,CYT)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19303. .19304,TYC)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19304. .19305,CYC)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19307. .19308,TMC)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19308. .19309,CCC)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19313. .19314,TMA)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19314. .19315,AWT)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19315. .19316,TCC)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19316. .19317,CYT)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19317. .19318,TWA)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19318. .19319,ATT)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19860,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20843,G)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21144,T)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21371,G)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22061. .22062,A)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22321,T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22470. .22471,A)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22472. .22473,A)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24990,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25284,A)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25934,G)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27509. .27510,TAT)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27513. .27514,TTT)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            27734. .27882
FT                   /*tag= at
FT                   /number= 2
FT   intron          27883. .41045
FT                   /*tag= au
FT                   /number= 2
FT   variation       replace(28608,G)
FT                   /*tag= av
```

```
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28627,C)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29709,T)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29788,T)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29878,T)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30294,T)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30896. .30897,T)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31815,G)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31999,TTT)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32026,T)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32146,A)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32184,S)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32214. .32215,T)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32215,C)
FT                   /*tag=  bi

  Query Match             2.2%;  Score 44.2;  DB 6;  Length 82938;
  Best Local Similarity   55.6%;  Pred. No. 4.3;
  Matches   85;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy       1623 ACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAA 1682
              |||||||    ||  ||  | | |||      | |||   || ||| || | |||| ||
Db      17088 ACAAAAAAGGGAAGGTAAATCAATAGGTTTCCTAGCACATTCTGCACATTTTAAATATTT 17029

Qy       1683 GAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTA 1742
              |  | |   | |||  ||||  || || | || ||  || ||   ||||  |||   |
Db      17028 GTCAGAAGGAACATTAGAAATACAGAAAAATTATTATATTAGTAAATTGGAAATACTTGG 16969

Qy       1743 AAAAGTATTTGAATTCTTTATGATGAGAACTAT 1775
              |||   | || ||   | ||| | | || |||
Db      16968 GAAATATTCTGTATAAATGATGGTAATAAGTAT 16936


RESULT 19
ABX13938
ID   ABX13938 standard; DNA; 3891 BP.
XX
AC   ABX13938;
XX
DT   03-MAR-2003  (first entry)
XX
DE   Clostridium botulinum, botulinum neurotoxin serotype A DNA.
XX
KW   Botulinum toxin; serotype A; neurotoxin; needleless syringe;
KW   involuntary muscle contraction; spasmodic dysphonia; laryngeal dystonia;
KW   oromandibular dysphonia; lingual dystonia; cervical dystonia;
KW   focal hand dystonia; blepharospasm; strabismus; hemifacial spasm;
KW   eyelid disorder; cerebral palsy; focal spasticity; spasmodic colitis;
KW   neurogenic bladder; anisumus; limb spasticity; tic; tremor; bruxism;
KW   anal fissure; achalasia; fibromyalgia; dysphagia; wrinkle; brow furrow;
KW   vaccine; motor function; cross-contamination; blood-borne pathogen;
KW   human immunodeficiency virus; HIV; hepatitis B; gene; ds.
XX
OS   Clostridium botulinum.
XX
PN   US2002107199-A1.
XX
PD   08-AUG-2002.
XX
PF   17-JAN-2002; 2002US-00051952.
XX
PR   05-DEC-2000; 2000US-00730237.
XX
PA   (ALLR ) ALLERGAN SALES INC.
XX
PI   Walker PS;
XX
DR   WPI; 2002-690769/74.
XX
PT   Treating a condition e.g., involuntary muscle contraction, spasmodic
PT   dysphonia, laryngeal dystonia, or wrinkles in an animal or human,
PT   comprises administering a Clostridium neurotoxin component using a
PT   needleless syringe.
XX
PS   Disclosure; Page 15-16; 33pp; English.
XX
CC   The invention describes a method of treating a condition in an animal or
CC   human comprising administering a Clostridium neurotoxin component (I) to
CC   the subject using a needleless syringe, or administering DNA encoding (I)
CC   to a cell of the subject in situ. The method is used for treating a
CC   condition e.g., involuntary muscle contraction or conditions such as
CC   spasmodic dysphonia, laryngeal dystonia, oromandibular dysphonia, lingual
CC   dystonia, cervical dystonia, focal hand dystonia, blepharospasm,
CC   strabismus, hemifacial spasm, eyelid disorder, cerebral palsy, focal
CC   spasticity, spasmodic colitis, neurogenic bladder, anisumus, limb
CC   spasticity, tics, tremors, bruxism, anal fissure, achalasia,
CC   fibromyalgia, dysphagia, wrinkles or brow furrows. The method is also
CC   useful for immunisation against a neurotoxin. By using a needleless
CC   injector, an optimal delivery to a specific tissue layer, for example the
CC   dermis layer is achieved. In a case where the delivery is to the dermis
CC   and not to muscle tissues, the treatment may not cause a loss of motor
CC   function in the area being treated. Also, the use of a needleless
CC   injector, improves clinical safety by eliminating the risk of infection
CC   from accidental injury with needles or from potential splash back of
CC   bodily fluids from liquid jet injectors, thus avoiding the possibilities
CC   of cross-contamination of blood-borne pathogens such as human
CC   immunodeficiency virus (HIV) and hepatitis B. The needleless injector
CC   also offers an optimal and specific delivery of drug particles to treat
CC   conditions with little pain or skin damage such as bruising or bleeding.
CC   This sequence represents the Clostridium botulinum neurotoxin serotype A
XX
SQ   Sequence 3891 BP; 1580 A; 386 C; 635 G; 1290 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 44;  DB 6;  Length 3891;
  Best Local Similarity   50.0%;  Pred. No. 1.3;
  Matches  110;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy       1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
              ||| |||| ||   |||  |  | |||||||  |||  || | ||     ||    |  |
Db       2919 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 2978

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
              |  | || ||      ||    |  ||   |||      ||| |||   |||| |  ||||
Db       2979 AAAACAAAGAGTAGTTTTTAAATAGAGTCAAATGATTAATATATCAGATTATATAAACAG 3038

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
              | ||||  | | ||| |  |  | | |  ||| |    ||| |  | |||  ||
Db       3039 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 3098

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
                || |    ||  | || | |||  ||| |     |||
Db       3099 AAGATTAATAGATGAAAAACCAATTTCAAATTTAGGTAAT 3138
```

RESULT 20
AAD42516
ID AAD42516 standard; DNA; 3891 BP.
XX
AC AAD42516;
XX
DT 15-NOV-2002 (first entry)
XX
DE Clostridium botulinum toxin A.
XX
KW Therapy; hyperhidrosis; drug; botulinum toxin; ds.
XX
OS Clostridium botulinum.
XX
PN US2002086036-A1.
XX
PD 04-JUL-2002.
XX
PF 05-DEC-2000; 2000US-00730237.
XX
PR 05-DEC-2000; 2000US-00730237.
XX
PA (ALLR) ALLERGAN SALES INC.
XX
PI Walker PS;
XX
DR WPI; 2002-618013/66.
XX
PT Treating hyperhidrosis in a mammal, comprises locally administering a
PT drug particle to an affected skin area without using a needle.
XX
PS Disclosure; Page 11-13; 29pp; English.
XX
CC The invention relates to a method for treating hyperhidrosis in a mammal,
CC which comprises locally administering a drug particle to an affected skin
CC area without using a needle. The method is useful for treating
CC hyperhidrosis. The present sequence is Clostridium botulinum toxin used
CC in the method of the invention
XX
SQ Sequence 3891 BP; 1580 A; 386 C; 635 G; 1290 T; 0 U; 0 Other;

Query Match 2.2%; Score 44; DB 6; Length 3891;
Best Local Similarity 50.0%; Pred. No. 1.3;
Matches 110; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
Db 2919 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 2978

Qy 1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
Db 2979 AAAACAAAGAGTAGTTTTTAAATAGAGTCAAATGATTAATATATCAGATTATATAAACAG 3038

Qy 1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
Db 3039 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 3098

Qy 1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db 3099 AAGATTAATAGATGAAAAACCAATTTCAAATTTAGGTAAT 3138

RESULT 21
ABZ35416/c
ID ABZ35416 standard; cDNA; 7093 BP.
XX
AC ABZ35416;
XX
DT 05-FEB-2003 (first entry)
XX
DE Human gene expression profile polynucleotide SEQ ID NO 527.
XX
KW Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;
KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;
KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;
KW gene expression; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200274979-A2.
XX
PD 26-SEP-2002.
XX
PF 20-MAR-2002; 2002WO-US008456.
XX
PR 20-MAR-2001; 2001US-0276947P.
XX
PA (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.
XX
PI Wan J, Wang Y;
XX
DR WPI; 2002-740862/80.
XX
PT New gene expression profile generated from primary, endothelial,
PT epithelial, and muscle cell types, useful for identifying disease
PT pathologies involving alterations of gene expression, e.g. cancer.
XX
PS Disclosure; Page 667-669; 850pp; English.
XX
CC The invention relates to a gene expression profile comprising one or more
CC genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type
CC is a coronary artery endothelium, umbilical artery or vein endothelium,
CC aortic endothelium, dermal microvascular endothelium, pulmonary artery
CC endothelium, myometrium microvascular endothelium, keratinocyte
CC epithelium, bronchial epithelium, mammary epithelium, prostate
CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,
CC small airway epithelium, renal epithelium, umbilical artery smooth
CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,
CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,
CC osteoblasts or prostate stromal cell. The gene expression profile is used
CC for determining the level of RNA expression for a sample, determining the
CC phenotype of a cell and distinguishing cell types. The gene or a protein
CC expression profile is useful in identifying disease pathologies involving
CC alterations of gene expression. The assessment of expression profiles may
CC provide meaningful information with respect to tumour type and stage,
CC treatment methods, and prognosis. The gene or protein expression profile
CC may also be used for creating microarrays. The microarray is useful for
CC genetic and physical mapping of genomes, DNA sequencing, genetic or
CC medical diagnosis, genotyping of organisms, confirming cell or tissue
CC identifications and in identifying promising antibiotics, antiviral or
CC antifungal agents
XX
SQ Sequence 7093 BP; 1999 A; 1487 C; 1453 G; 2154 T; 0 U; 0 Other;

Query Match 2.2%; Score 44; DB 6; Length 7093;
Best Local Similarity 63.0%; Pred. No. 1.7;
Matches 68; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 33 GTGAAAAATATATTAACGTGCATATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTC 92
Db 6619 GTCAAGTCTCAATGTTGGTAGATATATTAGTGTGGATTCAATTTCAAGTAACAGAAACCC 6560

Qy 93 AATTCAAACCAACGTAAGTCAAAAGGAAAATATATTGGCTCATGTAAC 140
Db 6559 AACTCAAAGTAACTTAAGCAAAAAAGGGTATTTATTGGTTCACATAAC 6512

RESULT 22
AAI99260/c
ID AAI99260 standard; DNA; 10126 BP.
XX
AC AAI99260;

XX
DT 07-JAN-2002 (first entry)
XX
DE Human excretory related polynucleotide SEQ ID NO 1024.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW excretory system; ds.
XX
OS Homo sapiens.
XX
PN WO200155313-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001323.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.

```
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465569/50.
XX
PT   Isolated nucleic acid molecule encoding excretory system antigen is used
PT   in preventing, treating or ameliorating a medical condition.
XX
PS   Example 2; SEQ ID NO 1024; 574pp + Sequence Listing; English.
XX
CC   The invention relates to novel excretory system related human
CC   polynucleotides (AAI98567-AAI99503) and the encoded proteins (AAM99594-
CC   AAM99913) useful for preventing, treating or ameliorating medical
CC   conditions e.g. by protein or gene therapy, espeically disorders related
CC   to the excretory system. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids, proteins,
CC   antibodies and (ant)agonists are useful in the diagnosis, treatment and
CC   prevention of: (a) cancer, e.g. breast and ovarian cancer and other
CC   cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal
CC   tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
CC   disease, allergies, autoimmune haemolytic anaemia, autoimmune
CC   thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
CC   rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders
CC   such as myocardial ischaemias; (d) wound healing; (e) neurological
CC   diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases
CC   such as viral, bacterial, fungal and parasitic infections. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 10126 BP; 2804 A; 2139 C; 2108 G; 3075 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 44;  DB 4;  Length 10126;
  Best Local Similarity   63.0%;  Pred. No. 2;
  Matches   68;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy         33 GTGAAAAATATATTAACGTGCATATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTC 92
              || ||   |  ||    ||  ||||| |   |     |  ||| ||||| |||||||| |
Db       9649 GTCAAGTCTCAATGTTGGTAGATATATTAGTGTGGATTCAATTTCAAGTAACAGAAACCC 9590

Qy         93 AATTCAAACCAACGTAAGTCAAAAGGAAAATATATTGGCTCATGTAAC 140
              || |||||  ||| ||||  |||| |   || |||||| |||  ||||
Db       9589 AACTCAAAGTAACTTAAGCAAAAAAGGGTATTTATTGGTTCACATAAC 9542


RESULT 23
AAI63610/c
ID   AAI63610 standard; DNA; 10126 BP.
XX
AC   AAI63610;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human kidney related polynucleotide SEQ ID NO 925.
XX
KW   Human; kidney antigen; immunosuppressive; antiarthritic; antirheumatic;
KW   antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW   nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW   opthalmological; antiallergic; hepatotropic; antidiabetic;
KW   antiinflammatory; antiulcer; vulnerary; anticonvulsant; antiparasitic;
KW   gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155323-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001343.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
```

PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488784/53.
XX
PT New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders.
XX
PS Disclosure; SEQ ID NO 925; 564pp + Sequence Listing; English.
XX
CC The invention relates to novel kidney related polynucleotides (AAI62971-
CC AAI63793) and the encoded polypeptides (AAM42417-AAM42691) collectively
CC known as kidney antigens and the use of such kidney antigens for
CC detecting disorders of the kidney, especially kidney cancer and kidney
CC cancer metastases. The polynucleotides and proteins are also useful for
CC preventing, treating or ameliorating medical conditions e.g. by protein
CC or gene therapy. The genes are isolated from a range of human tissues
CC disclosed in the specification. The nucleic acids, proteins, antibodies
CC and (ant)agonists are useful in the diagnosis, treatment and prevention
CC of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 10126 BP; 2804 A; 2139 C; 2108 G; 3075 T; 0 U; 0 Other;

Query Match 2.2%; Score 44; DB 5; Length 10126;
Best Local Similarity 63.0%; Pred. No. 2;
Matches 68; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 33 GTGAAAAATATATTAACGTGCATATAGTTCAGGACCATGGATTGCAAGTGACAGAAACTC 92
Db 9649 GTCAAGTCTCAATGTTGGTAGATATATTAGTGTGGATTCAATTTCAAGTAACAGAAACCC 9590

Qy 93 AATTCAAACCAACGTAAGTCAAAAGGAAAATATATTGGCTCATGTAAC 140
Db 9589 AACTCAAAGTAACTTAAGCAAAAAAGGGTATTTATTGGTTCACATAAC 9542

RESULT 24
ADR24206/c
ID ADR24206 standard; DNA; 5072 BP.
XX
AC ADR24206;
XX
DT 21-OCT-2004 (first entry)
XX
DE Breast cancer prognosis marker #67.
XX
KW ds; breast cancer; prognosis; gene expression; diagnosis.
XX
OS Homo sapiens.
XX
PN WO2004065545-A2.
XX
PD 05-AUG-2004.
XX
PF 15-JAN-2004; 2004WO-US001100.
XX
PR 15-JAN-2003; 2003US-00342887.
XX

PA (ROSE-) ROSETTA INPHARMATICS LLC.
PA (NECA-) NETHERLANDS CANCER INST.
XX
PI Van't Veer LJ, He Y;
XX
DR WPI; 2004-593473/57.
XX
PT Classifying a breast cancer patient according to prognosis comprises
PT determining the similarity between the level of expression of each of
PT five genes in a cell sample taken from patient, to control levels.
XX
PS Disclosure; SEQ ID NO 67; 226pp; English.
XX
CC The invention relates to a method of classifying a breast cancer patient
CC according to prognosis by determining the similarity between the level of
CC expression of each of five genes for which markers are listed in the
CC specification, in a cell sample taken from the breast cancer patient, to
CC control levels of expression for each respective five genes to obtain a
CC patient similarity value. The methods are useful for classifying a breast
CC cancer patient according to prognosis. Kits and computer program products
CC are useful for data analysis using the diagnostic, prognostic and
CC statistical methods of the invention. This sequence corresponds to a
CC marker used in the method of the invention.
XX
SQ Sequence 5072 BP; 1309 A; 1136 C; 1406 G; 1221 T; 0 U; 0 Other;

Query Match 2.2%; Score 43.6; DB 13; Length 5072;
Best Local Similarity 50.8%; Pred. No. 1.9;
Matches 130; Conservative 0; Mismatches 124; Indels 2; Gaps 1;

Qy 1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
Db 2028 AAACTTAAAAAAAAAAAAAAGAACAGGACTAGAAAACTTAATTTTTTTTGACATATCTAAA 1969

Qy 1730 TGAGAATCTATTAAAA--AGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGT 1787
Db 1968 AAAGGGGACAAGAACAGTAGCATTTGTTCTCCCCATGGGGACAGGAATGAAGAGGGAGGA 1909

Qy 1788 GGACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAG 1847
Db 1908 GGGAAGAGGGACTTGATTTCTTCTGTTCACAAGAGAATGGTTGGAGGGGATTCTTCAGAG 1849

Qy 1848 TCTTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGAT 1907
Db 1848 TTTTTCAGGGGATGGGACCCAAATGTAGTGCCCAGACTGGTCCTCAATGATTTGTTTGAT 1789

Qy 1908 TTTGTTTTAGACTTTT 1923
Db 1788 TTTGTTATAAATCTCT 1773

RESULT 25
ACC85075/c
ID ACC85075 standard; DNA; 5073 BP.
XX
AC ACC85075;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human MBCAT polypeptide encoding DNA.
XX
KW MBCAT; beta-catenin; cytostatic; gene therapy; cancer; human; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 317. .1855
FT /*tag= a
FT /product= "MBCAT"
XX
PN WO2003052068-A2.
XX
PD 26-JUN-2003.
XX
PF 12-DEC-2002; 2002WO-US039796.
XX
PR 13-DEC-2001; 2001US-0340213P.
PR 13-DEC-2001; 2001US-0340314P.
PR 13-DEC-2001; 2001US-0340322P.
PR 15-FEB-2002; 2002US-0357502P.
XX
PA (EXEL-) EXELIXIS INC.
XX
PI Costa MA, Gendreau SB, Dora EG, Nicoll M;
XX
DR WPI; 2003-533010/50.
DR P-PSDB; ABR82224.
XX
PT Identifying a candidate beta-catenin pathway modulating agent for
PT diagnosing or treating cancer by detecting a test agent-biased activity
PT of the assay system comprising a purified MBCAT polypeptide or nucleic
PT acid.
XX
PS Example; Page 57-59; 81pp; English.
XX
CC The invention relates to genes that modify beta-catenin pathway and to
CC the identification of human MBCAT (modifiers of beta-catenin) genes. The
CC MBCAT polypeptides are therapeutic targets for disorders associated with
CC beta-catenin function and are useful for manufacturing a medicament for
CC diagnosing or treating breast, colon, lung or ovary cancer. The present
CC sequence represents a human MBCAT polypeptide encoding DNA
XX
SQ Sequence 5073 BP; 1309 A; 1136 C; 1408 G; 1220 T; 0 U; 0 Other;

Query Match 2.2%; Score 43.6; DB 9; Length 5073;
Best Local Similarity 50.8%; Pred. No. 1.9;
Matches 130; Conservative 0; Mismatches 124; Indels 2; Gaps 1;

Qy 1670 AATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCAT 1729
Db 2029 AAACTTAAAAAAAAAAAAAAGAACAGGACTAGAAAACTTAATTTTTTTTGACATATCTAAA 1970

Qy 1730 TGAGAATCTATTAAAA--AGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGT 1787
Db 1969 AAAGGGGACAAGAACAGTAGCATTTGTTCTCCCCATGGGGACAGGAATGAAGAGGGAGGA 1910

Qy 1788 GGACAGTGGTGAGCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAG 1847
Db 1909 GGGAAGAGGGACTTGATTTCTTCTGTTCACAAGAGAATGGTTGGAGGGGATTCTTCAGAG 1850

Qy 1848 TCTTATATGGTACAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGAT 1907
Db 1849 TTTTTCAGGGGATGGGACCCAAATGTAGTGCCCAGACTGGTCCTCAATGATTTGTTTGAT 1790

Qy 1908 TTTGTTTTAGACTTTT 1923
Db 1789 TTTGTTATAAATCTCT 1774

RESULT 26
AAS46304/c
ID AAS46304 standard; DNA; 10369 BP.
XX
AC AAS46304;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #26.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.

```
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 26; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Sequences with even
CC   numbered Seq ID numbers are the complementary sequence of the
CC   corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC   535, except for those whose partner sequence is missing). Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 10369 BP; 2628 A; 319 C; 2617 G; 4805 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 43.6;  DB 4;  Length 10369;
  Best Local Similarity   51.5%;  Pred. No. 2.6;
  Matches  100;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy       1581 ATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTAT 1640
             ||| || |  | |||| |||   ||  ||| | | ||  || ||  ||||    ||
Db       8333 ATTATACTTCAAATTTATAATACTAAAAACATAAACAAACTAAAAATTACTAACCACTTA 8274

Qy       1641 GTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGA 1700
               ||  | || ||         ||||   ||| | |||||  | ||| |  ||| |
Db       8273 TCAAAAACAATAAATCCAACTTTTTATACAAAAAAAAAAATACAAATATTAACTAAACAT 8214

Qy       1701 AAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTT 1760
             ||    | |||  | |     ||  |  | || || |||  | |||  |   | ||  ||
Db       8213 AATAATACATACCTATAATCCCAACTACTCGAAAAACTAAAACAAAATTACTTAAACCTA 8154

Qy       1761 TATGATGAGAACTA 1774
               |  | ||  |||||
Db       8153 AAAAACGAAAACTA 8140


RESULT 27
ABL32393/c
ID   ABL32393 standard; DNA; 10369 BP.
XX
AC   ABL32393;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 366.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 366; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 10369 BP; 2628 A; 319 C; 2617 G; 4805 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 43.6;  DB 6;  Length 10369;
  Best Local Similarity   51.5%;  Pred. No. 2.6;
  Matches  100;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy       1581 ATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTAT 1640
             ||| || |  | |||| |||   ||  ||| | | ||  || ||  ||||    ||
Db       8333 ATTATACTTCAAATTTATAATACTAAAAACATAAACAAACTAAAAATTACTAACCACTTA 8274

Qy       1641 GTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGA 1700
               ||  | || ||         ||||   ||| | |||||  | ||| |  ||| |
Db       8273 TCAAAAACAATAAATCCAACTTTTTATACAAAAAAAAAAATACAAATATTAACTAAACAT 8214

Qy       1701 AAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTT 1760
             ||    | |||  | |     ||  |  | || || |||  | |||  |   | ||  ||
Db       8213 AATAATACATACCTATAATCCCAACTACTCGAAAAACTAAAACAAAATTACTTAAACCTA 8154

Qy       1761 TATGATGAGAACTA 1774
               |  | ||  |||||
Db       8153 AAAAACGAAAACTA 8140


RESULT 28
```

```
ADR85280/c
ID   ADR85280 standard; DNA; 1212 BP.
XX
AC   ADR85280;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Aspergillus fumigatus essential gene with introns #504.
XX
KW   Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;
KW   drug screening; ds.
XX
OS   Aspergillus fumigatus.
XX
PN   WO2004067709-A2.
XX
PD   12-AUG-2004.
XX
PF   16-JAN-2004; 2004WO-US001099.
XX
PR   17-JAN-2003; 2003US-0441281P.
PR   13-JUN-2003; 2003US-0478196P.
XX
PA   (ELIT-) ELITRA PHARM INC.
PA   (ELIT-) ELITRA CANADA LTD.
XX
PI   Jiang B,  Hu W,  Lemieux S,  Roemer T;
XX
DR   WPI; 2004-594200/57.
DR   P-PSDB; ADR86454.
XX
PT   New purified or isolated Aspergillus fumigatus nucleic acid molecule
PT   encoding a gene product, useful for diagnosing and/or treating invasive
PT   fungal infections, such as Farmer's lung disease.
XX
PS   Claim 3; SEQ ID NO 1504; 164pp; English.
XX
CC   The present invention relates to Aspergillus fumigatus genes that are
CC   essential and are potential targets for drug screening. The methods and
CC   compositions of the present invention are useful for diagnosing and/or
CC   treating invasive Aspergillus fumigatus infection, including the allergic
CC   forms of the disease, such as Farmer's lung disease. They can also be
CC   used in various drug discovery purposes, such as expression of the
CC   recombinant protein, hybridization assay and construction of nucleic acid
CC   arrays. The present sequence represents an Aspergillus fumigatus
CC   essential gene sequence containing all introns, used during diagnosis and
CC   drug development in the invention. These genes share a high degree of
CC   sequence conservation with known essential genes of candida albicans. The
CC   sequence data for this patent is not represented in the printed
CC   specification, but was obtained in electronic format from WIPO.
XX
SQ   Sequence 1212 BP; 311 A; 328 C; 335 G; 238 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 43.4;  DB 13;  Length 1212;
  Best Local Similarity   50.7%;  Pred. No. 1.2;
  Matches  104;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        163 GGAAGGGGCTTTGGGAACAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCC 222
              |||  |   ||||||  ||||||   ||| |  | |||||  | ||  ||   | |     |
Db        756 GGAGAGATATTTGGGGACAAGATCCTTGCTGGCCAATTCGACGACCACCTTGCTCTCGGC 697

Qy        223 AGGATGGGTCACCTTCCTGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGG 282
               |||||      |   | | |   ||  || | || ||||  ||  | |  ||  ||||||
Db        696 TGGATGCTGTTCACCCATATATGTGGTGTAGCGGAAGCGCAGGATGTGAGTTTCGGGAGG 637

Qy        283 AAAGAGTGCATGTTAGGATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATC 342
              ||      | |||  ||       |||||  |  |  | ||                || |||
Db        636 AAGAGTAAAAGGTTTGGCGAGTTCTAGGGATTGGGGACCAGATTAGCATTGAACCATAGT 577

Qy        343 ATGCTAAAAATGGTTTTTTTTGATG 367
                |||||||||  | |   ||  ||
Db        576 CGCCTAAAAATGTTGTAACTTACTG 552
```

```
RESULT 29
ADZ70960/c
ID   ADZ70960 standard; DNA; 1236 BP.
XX
AC   ADZ70960;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Human chromosome 1 contig DNA SEQ ID NO 84.
XX
KW   ds; matrix attachment region; MAR; protein production.
XX
OS   Homo sapiens.
XX
PN   WO2005040377-A2.
XX
PD   06-MAY-2005.
XX
PF   22-OCT-2004; 2004WO-EP011974.
XX
PR   24-OCT-2003; 2003US-0513574P.
PR   06-FEB-2004; 2004EP-00002722.
XX
PA   (SELE-) SELEXIS SA.
XX
PI   Mermod N,  Girod PA,  Bucher P,  Nguyen D,  Calabrese D,  Saugy D;
PI   Puttini S;
XX
DR   WPI; 2005-333507/34.
XX
PT   New purified and isolated DNA sequence having protein production
PT   increasing activity comprises a bent DNA element and a binding site for a
PT   DNA binding protein, useful for increasing protein production activity in
PT   eukaryotic host cell.
XX
PS   Disclosure; SEQ ID NO 84; 282pp; English.
XX
CC   The invention relates to a purified and isolated DNA sequence having
CC   protein production increasing activity comprising at least one bent DNA
CC   element, and at least one binding site for a DNA binding protein. The
CC   purified and isolated DNA sequence comprising a first and a second
CC   isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC   MAR nucleotide sequence selected from a purified and isolated DNA
CC   sequence above, a purified and isolated MAR DNA above, a purified and
CC   isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC   sequence complementary to it, its molecular chimera, or its combinations
CC   and variants, is useful for increasing protein production activity in a
CC   eukaryotic host cell. The present sequence represents a human chromosome
CC   1 contig DNA.
XX
SQ   Sequence 1236 BP; 506 A; 28 C; 126 G; 576 T; 0 U; 0 Other;

  Query Match             2.2%;  Score 43.4;  DB 14;  Length 1236;
  Best Local Similarity   45.1%;  Pred. No. 1.2;
  Matches  161;  Conservative    0;  Mismatches  196;  Indels    0;  Gaps    0;

Qy       1408 TAATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACTAAG 1467
              ||||||||       ||| | |  |||  |      |    | || | ||||||  |||||
Db        397 TAATACATCATATATATTATATATGTTATATAATACATCATATATATTATATATACTAAA 338

Qy       1468 GCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCAC 1527
                 | || |||  | |        | |   || | |||||  |    |
Db        337 TAATACATCATATATAATATATACTATATAATACATATAATACATTACATATACTATATA 278

Qy       1528 AGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTC 1587
              | | | |||  | |   || |  || ||    || |||| | |      ||  || |
Db        277 ATATATAATACATTACATATACTATACATAATATGTAATACATTATATAGTATATATACT 218

Qy       1588 TGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTA 1647
                | |   || |||   |    ||    ||    | | | |||| |     | ||
Db        217 ATATAATACATTATATATACTATATATAATACATTATATATACTATATATACTATATAAT 158
```

```
Qy       1648 GAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAA 1707
Db        157 ATATAATACATTATATATACTATATATACTATATAATATATAATACATTATATAAATAAT 98

Qy       1708 AATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATG 1764
Db         97 AATATATAATATATAATATATAATAATATATAATATATATTATATATAATACATATG 41
```

RESULT 30
ADR84693/c
ID ADR84693 standard; DNA; 3212 BP.
XX
AC ADR84693;
XX
DT 04-NOV-2004 (first entry)
XX
DE Aspergillus fumigatus essential gene genomic sequence #504.
XX
KW Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;
KW drug screening; ds.
XX
OS Aspergillus fumigatus.
XX
PN WO2004067709-A2.
XX
PD 12-AUG-2004.
XX
PF 16-JAN-2004; 2004WO-US001099.
XX
PR 17-JAN-2003; 2003US-0441281P.
PR 13-JUN-2003; 2003US-0478196P.
XX
PA (ELIT-) ELITRA PHARM INC.
PA (ELIT-) ELITRA CANADA LTD.
XX
PI Jiang B, Hu W, Lemieux S, Roemer T;
XX
DR WPI; 2004-594200/57.
DR P-PSDB; ADR86454.
XX
PT New purified or isolated Aspergillus fumigatus nucleic acid molecule
PT encoding a gene product, useful for diagnosing and/or treating invasive
PT fungal infections, such as Farmer's lung disease.
XX
PS Claim 3; SEQ ID NO 504; 164pp; English.
XX
CC The present invention relates to Aspergillus fumigatus genes that are
CC essential and are potential targets for drug screening. The methods and
CC compositions of the present invention are useful for diagnosing and/or
CC treating invasive Aspergillus fumigatus infection, including the allergic
CC forms of the disease, such as Farmer's lung disease. They can also be
CC used in various drug discovery purposes, such as expression of the
CC recombinant protein, hybridization assay and construction of nucleic acid
CC arrays. The present sequence represents an Aspergillus fumigatus
CC essential gene full length genomic sequence, used during diagnosis and
CC drug development in the invention. These genes share a high degree of
CC sequence conservation with known essential genes of candida albicans. The
CC sequence data for this patent is not represented in the printed
CC specification, but was obtained in electronic format from WIPO.
XX
SQ Sequence 3212 BP; 870 A; 810 C; 762 G; 770 T; 0 U; 0 Other;

```
  Query Match             2.2%;  Score 43.4;  DB 13;  Length 3212;
  Best Local Similarity   50.7%;  Pred. No. 1.8;
  Matches  104;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        163 GGAAGGGGCTTTGGGAACAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCC 222
Db       1756 GGAGAGATATTTGGGGACAAGATCCTTGCTGGCCAATTCGACGACCACCTTGCTCTCGGC 1697

Qy        223 AGGATGGGTCACCTTCCTGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGG 282
Db       1696 TGGATGCTGTTCACCCATATATGTGGTGTAGCGGAAGCGCAGGATGTGAGTTTCGGGAGG 1637

Qy        283 AAAGAGTGCATGTTAGGATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATC 342
Db       1636 AAGAGTAAAAGGTTTGGCGAGTTCTAGGGATTGGGGACCAGATTAGCATTGAACCATAGT 1577

Qy        343 ATGCTAAAAATGGTTTTTTTTGATG 367
Db       1576 CGCCTAAAAATGTTGTAACTTACTG 1552
```

RESULT 31
AAL06925/c
ID AAL06925 standard; DNA; 32216 BP.
XX
AC AAL06925;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 9613.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 9613; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 32216 BP; 8496 A; 6430 C; 6479 G; 10811 T; 0 U; 0 Other;

Query Match 2.1%; Score 43; DB 4; Length 32216;
Best Local Similarity 55.8%; Pred. No. 6;
Matches 82; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy      1653 ACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATAT 1712
             ||| | ||   || |     | ||||||||  |||  |||  |  | ||  | ||||||
Db     23479 ACAGGGTGAGATTCCATCTCAAAAAAATAAATAAATAAAATAAAATTAATCTGAAAATAC 23420

Qy      1713 TTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAAC 1772
             |  || ||    | |||    ||    ||| ||| ||||| || | | || | |   ||
Db     23419 TGATTAAAGAGAGACATCTCTAACTATTTAGAAAATATTTTAAATTTGTAGGTTTTTAAT 23360

Qy      1773 TATCTTGACTCAAGTGGACAGTGGTGA 1799
             | |  |  || ||    | | |||||||
Db     23359 TTTACTACCTGAATCAAAAATTGGTGA 23333
```

RESULT 32
ABA07741/c
ID ABA07741 standard; DNA; 32216 BP.
XX
AC ABA07741;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 536.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;

KW neurological disease; infection; human; secreted protein; ds.
XX
OS Homo sapiens.
XX
PN WO200155325-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001345.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

```
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488786/53.
XX
PT   New isolated ovarian and/or breast cancer related nucleic acids and
PT   polypeptides, useful for diagnosing, treating and/or preventing human
PT   diseases and disorders, particularly ovarian and/or breast cancer.
XX
PS   Disclosure; SEQ ID NO 536; 577pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA07454-ABA08224) and proteins
CC   (ABB10743-ABB10980) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 32216 BP; 8496 A; 6430 C; 6479 G; 10811 T; 0 U; 0 Other;

  Query Match             2.1%;  Score 43;  DB 4;  Length 32216;
  Best Local Similarity   55.8%;  Pred. No. 6;
  Matches   82;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy       1653 ACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATAT 1712
             ||| | ||   || |     | ||||||||  |||  |||  |  | ||  | ||||||
Db      23479 ACAGGGTGAGATTCCATCTCAAAAAAATAAATAAATAAAATAAAATTAATCTGAAAATAC 23420

Qy       1713 TTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAAC 1772
              |  || ||    | |||    ||    ||| ||| ||||| || | | || | |   ||
Db      23419 TGATTAAAGAGAGACATCTCTAACTATTTAGAAAATATTTTAAATTTGTAGGTTTTTAAT 23360

Qy       1773 TATCTTGACTCAAGTGGACAGTGGTGA 1799
              | |  |  || ||    | | ||||||
Db      23359 TTTACTACCTGAATCAAAAATTGGTGA 23333


RESULT 33
ADF51132
ID   ADF51132 standard; DNA; 243428 BP.
XX
AC   ADF51132;
XX
DT   12-FEB-2004  (first entry)
XX
DE   Human P-Rex1 genomic DNA sequence.
XX
KW   human; P-Rex1; Rac; guanine-nucleotide exchange factor; GEF; GTPase;
KW   inflammation; metastasis; septic shock; neurodegeneration;
KW   atherosclerosis; antiinflammatory; cytostatic; antibacterial;
KW   immunosuppressive; neuroprotective; antiarteriosclerotic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003080664-A1.
XX
PD   02-OCT-2003.
XX
PF   21-MAR-2003; 2003WO-GB001238.
XX
PR   21-MAR-2002; 2002GB-00006684.
XX
PA   (BABR-) BABRAHAM INST.
XX
PI   Stephens L,  Hawkins PT;
XX
DR   WPI; 2004-011515/01.
DR   P-PSDB; ADF51119.
XX
PT   New isolated P-Rex1 protein or its derivative useful for discovering
PT   drugs capable of reducing or inhibiting inflammation, metastasis, septic
PT   shock, neurodegeneration or atherosclerosis, or for identifying P-Rex1
PT   modulators.
XX
PS   Disclosure; SEQ ID NO 14; 198pp; English.
XX
CC   This invention relates to a novel protein useful as an anti-inflammatory
CC   target. Specifically, it refers to the guanine-nucleotide exchange factor
CC   (GEF) named P-Rex1, which has also been identified as a
CC   phosphatidylinositol(3,4,5)P3-sensitive activator of Rac (a monomeric
CC   GTPase). Accordingly, P-Rex1 can be described as having Rac-GEF activity
CC   and is adapted to function downstream of activation of heterotrimeric G
CC   proteins in neutrophils. The present invention describes this protein as
CC   a useful target for drug discovery or for discovery of a drug capable of
CC   reducing or inhibiting inflammation, metastasis, septic shock,
CC   neurodegeneration or atherosclerosis. As such, P-Rex1 can have various
CC   activities including antiinflammatory, cytostatic, antibacterial,
CC   immunosuppressive, neuroprotective and antiarteriosclerotic. Furthermore,
CC   the protein or its mutant, the nucleic acid or appropriate antibody may
CC   be used in a screening assay to identify a modulator of P-Rex1 binding,
CC   activity or expression. This polynucleotide is the human P-Rex1 genomic
CC   DNA sequence of the invention.
XX
SQ   Sequence 243428 BP; 65880 A; 63219 C; 59010 G; 55319 T; 0 U; 0 Other;

  Query Match             2.1%;  Score 43;  DB 12;  Length 243428;
  Best Local Similarity   59.3%;  Pred. No. 14;
  Matches   73;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy         75 TGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATATTGGCTCA 134
              ||||||||||||||  ||| ||||||||    | |  |||     |||||||||| |||
Db     238428 TGCAAGTGACAGAAATCCAACTCAAACCAGTTCAGGGAAAATACGGAATATATTGGGTCA 238487

Qy        135 TGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAACAAGAGAATTGTTCT 194
              | ||| | | ||  |    |||||  |   | |       |    ||  |  |  |  |||
Db     238488 TCTAAGCCTGTCTAACTTCGGGCATGGCTGGATCCAGGTGCTCAAAGATTGCTTCTTTCC 238547

Qy        195 CAA 197
              |||
Db     238548 CAA 238550


RESULT 34
ABL32695/c
ID   ABL32695 standard; DNA; 9963 BP.
XX
AC   ABL32695;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 668.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 668; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 9963 BP; 2584 A; 180 C; 2281 G; 4918 T; 0 U; 0 Other;

  Query Match             2.1%;  Score 42.8;  DB 6;  Length 9963;
  Best Local Similarity   52.2%;  Pred. No. 4.1;
  Matches   95;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy       1596 TATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACA 1655
             ||| |  ||  ||  | ||  ||  |  ||||| |  ||||||| |||     | || |
Db       7988 TATTCAAAAAAATAAAATAAAAAAAAACTAAAAAATTACATCTATATAAAATTTATAAAA 7929

Qy       1656 AGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTG 1715
              |          |  | ||| |||||| | ||| |||  |  | ||| | |||| | | |
Db       7928 ATAAAAAAACCATATACATACAAATAAAATAAAAAAAAAACATAAAAATAAAAAAACTAC 7869

Qy       1716 TTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTAT 1775
               ||||| |  | ||  | ||||| || |      |  |   || |  || || |  |  ||
Db       7868 ATTAATAAAATTATAAACAATCTCTTTATTTTAAAATTCTTTTTAAATAATTATTATAAT 7809

Qy       1776 CT 1777
              |
Db       7808 AT 7807


RESULT 35
AEB42401_32/c
Continuation (33 of 34) of AEB42401 from base 3200001 (L. pneumophila DNA SEQ ID NO 6733
WP Sequence split into 34 fragments  LOCUS AEB42401 Accession Aeb42401
WP      Fragment Name      Begin      End
WP      AEB42401_00            1    110000
WP      AEB42401_01       100001    210000
WP      AEB42401_02       200001    310000
WP      AEB42401_03       300001    410000
WP      AEB42401_04       400001    510000
WP      AEB42401_05       500001    610000
WP      AEB42401_06       600001    710000
WP      AEB42401_07       700001    810000
WP      AEB42401_08       800001    910000
WP      AEB42401_09       900001   1010000
WP      AEB42401_10      1000001   1110000
WP      AEB42401_11      1100001   1210000
WP      AEB42401_12      1200001   1310000
WP      AEB42401_13      1300001   1410000
WP      AEB42401_14      1400001   1510000
WP      AEB42401_15      1500001   1610000
WP      AEB42401_16      1600001   1710000
WP      AEB42401_17      1700001   1810000
WP      AEB42401_18      1800001   1910000
WP      AEB42401_19      1900001   2010000
WP      AEB42401_20      2000001   2110000
WP      AEB42401_21      2100001   2210000
WP      AEB42401_22      2200001   2310000
WP      AEB42401_23      2300001   2410000
WP      AEB42401_24      2400001   2510000
WP      AEB42401_25      2500001   2610000
WP      AEB42401_26      2600001   2710000
WP      AEB42401_27      2700001   2810000
WP      AEB42401_28      2800001   2910000
WP      AEB42401_29      2900001   3010000
WP      AEB42401_30      3000001   3110000
WP      AEB42401_31      3100001   3210000
WP      AEB42401_32      3200001   3310000
WP      AEB42401_33      3300001   3345687

  Query Match             2.1%;  Score 42.8;  DB 14;  Length 110000;
  Best Local Similarity   46.9%;  Pred. No. 11;
  Matches  134;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy       1531 TCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGC 1590
              ||| |||||   ||  | |    | | |   | |     |  | |   |||| ||  |
Db      82357 TCTATTATGTCTGGAGATTCTATTTTGGACAACATTGATTTTTTGTTGTCATTAATCGCC 82298

Qy       1591 TCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAA 1650
              |  | ||||  ||  ||    |    |  | || |  | ||| ||   ||   |||| ||
Db      82297 TTTCATATCCAAAACAATGGGTATCTTTTGAAAGACGAGCTAAAAAAGATTGGATTAAAA 82238

Qy       1651 GAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAAT 1710
                |  |   ||   | |||  | | | ||  | | |  |  || |   || |    | |||
Db      82237 ATAGTATTGGGGAGTACTCGGTTTCAGAAATAAATACTGTCACAAATGGACAAATATAAT 82178

Qy       1711 ATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGA 1770
                ||  | | |  | ||||| | |  | | |  |  |    | |||| ||  ||  | |
Db      82177 TTTCCTGTTAATATGTCATAGTGTTTTTTTAGAGGATCTATCAAATTTTTAGGGAGTAAA 82118

Qy       1771 ACTATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTGTGGT 1816
                |  |   ||  | |      | || |||||| ||   || | |
Db      82117 AGCAGTATGTATTTAAATTCACGCTGTAAGCTTTATGCTTTGGGTT 82072


RESULT 36
ABL33091/c
ID   ABL33091 standard; DNA; 16724 BP.
XX
AC   ABL33091;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1064.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
```

PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1064; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 16724 BP; 5336 A; 237 C; 3558 G; 7593 T; 0 U; 0 Other;

Query Match 2.1%; Score 42.6; DB 6; Length 16724;
Best Local Similarity 50.2%; Pred. No. 5.8;
Matches 135; Conservative 0; Mismatches 129; Indels 5; Gaps 1;

Qy 1517 CTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCT 1576
Db 16680 CCTTACAAAACAATAAAAATATATTAATTATTATTAAATTCGTAAACAATAAAATTAATT 16621

Qy 1577 GTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAAT 1636
Db 16620 CCTAAAATCTATAAACTCTAAAAATACAAAAACAATAAAATATAATTCCTAAAAAAAATC 16561

Qy 1637 CTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCAT 1696
Db 16560 TTTCTTAAAAACAAACTAACGCCTAATACAAACAATAAAAAAAAAATAAAAAACAACCAC 16501

Qy 1697 GTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db 16500 TTTAAAAACATTATATATATTTATATATAACA-----AAAAAATAAAAACTTATAAAAAA 16446

Qy 1757 TCTTTATGATGAGAACTATCTTGACTCAA 1785
Db 16445 TAATTACAAACAAAACTAATATTACCCAA 16417

RESULT 37
ABL34537/c
ID ABL34537 standard; DNA; 16724 BP.
XX
AC ABL34537;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human metastasis associated gene SEQ ID NO: 90.
XX
KW Metastasis associated gene; cytostatic; gene therapy; cancer;
KW cytosine methylation; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200177376-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003970.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-010922/01.
XX
PT New nucleic acid derived from chemically treated metastasis genes, useful
PT for diagnosis of cancers by analysis of cytosine methylation, also for
PT treatment.
XX
PS Claim 1; SEQ ID NO 90; 23pp + Sequence Listing; English.
XX
CC The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention
XX
SQ Sequence 16724 BP; 5336 A; 237 C; 3558 G; 7593 T; 0 U; 0 Other;

Query Match 2.1%; Score 42.6; DB 6; Length 16724;
Best Local Similarity 50.2%; Pred. No. 5.8;
Matches 135; Conservative 0; Mismatches 129; Indels 5; Gaps 1;

Qy 1517 CTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCT 1576
Db 16680 CCTTACAAAACAATAAAAATATATTAATTATTATTAAATTCGTAAACAATAAAATTAATT 16621

Qy 1577 GTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAAT 1636
Db 16620 CCTAAAATCTATAAACTCTAAAAATACAAAAACAATAAAATATAATTCCTAAAAAAAATC 16561

Qy 1637 CTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCAT 1696
Db 16560 TTTCTTAAAAACAAACTAACGCCTAATACAAACAATAAAAAAAAAATAAAAAACAACCAC 16501

Qy 1697 GTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db 16500 TTTAAAAACATTATATATATTTATATATAACA-----AAAAAATAAAAACTTATAAAAAA 16446

Qy 1757 TCTTTATGATGAGAACTATCTTGACTCAA 1785
Db 16445 TAATTACAAACAAAACTAATATTACCCAA 16417

RESULT 38
ABL70260/c
ID ABL70260 standard; DNA; 16724 BP.
XX
AC ABL70260;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence complementary to#75.
XX
KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW tumour; cytostatic; ds.
XX
OS Unidentified.
XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007471.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX

```
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154758/20.
XX
PT   Nucleic acid, useful for diagnosis and therapy of diseases associated
PT   with cell signaling e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with cell signaling.
XX
PS   Claim 1; SEQ ID NO 150; 24pp + Sequence Listing; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of at least
CC   18 bases of a segment of chemically pretreated DNA of genes associated
CC   with cell signalling. The activity of the modified sequences of the
CC   invention may be described as cytostatic. The object of the invention is
CC   to provide the chemically modified DNA of genes associated with cell
CC   signalling, as well as oligonucleotides and/or PNA-oligomers for
CC   detecting cytosine methylations, as well as a method which is
CC   particularly suitable for the diagnosis and/or therapy of genetic and
CC   epigenetic parameters of genes associated with cell signalling. The
CC   chemically modified DNA provided by the invention is useful for diagnosis
CC   and therapy of diseases such as solid tumours and cancer. The sequences
CC   given in records ABL70111-ABL70626 represent chemically pre-treated
CC   genomic DNA's of genes associated with cell signalling. Note: The
CC   sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office
XX
SQ   Sequence 16724 BP; 5336 A; 237 C; 3558 G; 7593 T; 0 U; 0 Other;

  Query Match             2.1%;  Score 42.6;  DB 6;  Length 16724;
  Best Local Similarity   50.2%;  Pred. No. 5.8;
  Matches  135;  Conservative    0;  Mismatches  129;  Indels    5;  Gaps    1;

Qy       1517 CTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCT 1576
             | || |   |||       |||| | ||  |||  | |  | |     |||| | | |  |
Db      16680 CCTTACAAAACAATAAAAATATATTAATTATTATTAAATTCGTAAACAATAAAATTAATT 16621

Qy       1577 GTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAAT 1636
                |  |||  |   | || |  |||   ||   || |  ||     |  |||  | | 
Db      16620 CCTAAAATCTATAAACTCTAAAAATACAAAAACAATAAAATATAATTCCTAAAAAAAATC 16561

Qy       1637 CTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCAT 1696
                |   |||  | ||    | ||    |     ||||| ||||| || |||||  |||||
Db      16560 TTTCTTAAAAACAAACTAACGCCTAATACAAACAATAAAAAAAAAATAAAAAACAACCAC 16501

Qy       1697 GTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
                | |||  ||  |||| | ||||   |   ||     ||   || ||||  |||   ||
Db      16500 TTTAAAAACATTATATATATTTATATATAACA-----AAAAATAAAAACTTATAAAAAA 16446

Qy       1757 TCTTTATGATGAGAACTATCTTGACTCAA 1785
              |  |||  |  | |||||   | || |||
Db      16445 TAATTACAAACAAAACTAATATTACCCAA 16417


RESULT 39
ADS99798/c
ID   ADS99798 standard; DNA; 16724 BP.
XX
AC   ADS99798;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Complement of bisulphite treated metastasis-associated human gene #45.
XX
KW   Human; ds; gene; Bisulphite; metastasis; cancer; cytostatic;
KW   DNA methylation; matrix-assisted laser desorption/ionisation; MALDI;
KW   electrospray; mass spectrometry; CpG dinucleotide; solid tumour.
XX
OS   Homo sapiens.
XX
PN   US2003148327-A1.
XX
PD   07-AUG-2003.
XX
PF   21-JAN-2003; 2003US-00240485.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
PR   06-APR-2001; 2001WO-EP003970.
XX
PA   (OLEK/) OLEK A.
PA   (PIEP/) PIEPENBROCK C.
PA   (BERL/) BERLIN K.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-010922/01.
XX
PT   New nucleic acid derived from chemically treated metastasis genes, useful
PT   for diagnosis of cancers by analysis of cytosine methylation, also for
PT   treatment.
XX
PS   Claim 1; SEQ ID NO 90; 9pp; English.
XX
CC   The invention relates to a nucleic acid comprising at least 18 bases from
CC   a segment of the chemically pretreated DNA of genes associated with
CC   metastasis, i.e. any of ADS99709-ADS99906 human genomic sequences or any
CC   of the 19 sequences appearing as ADS99911-ADS99929. SEQ ID 2,4,6 etc are
CC   the complements of SEQ ID 1,3,5, etc. Also included are an oligomer
CC   (particularly an oligonucleotide or peptide nucleic acid) comprising at
CC   least one base sequence of at least 9 bases which hybridises to (or is
CC   identical with) the sequences referred to above, producing an array of
CC   the oligomers on a carrier, obtaining genetic and/or epigenetic
CC   parameters for diagnosis and/or therapy of diseases (or predisposition to
CC   them) by analysis of cytosine methylation and a kit comprising a
CC   bisulphite (disulphite or hydrogen sulphite) and the oligomers. In the
CC   method of above 5-unmethylated cytosines in a genomic DNA sample are
CC   converted chemically to uracil, or another base with hybridisation
CC   properties different from those of cytosine, then fragments of the
CC   treated DNA amplified (particularly by polymerase chain reaction) using
CC   the oligomers and a polymerase (preferably heat stable) to produce
CC   labelled amplicons. These are tested for hybridisation to an array of
CC   oligomers and any hybridisation detected. The amplicons are labelled with
CC   fluorescent or radioactive markers, or with a detachable mass marker to
CC   allow their detection by mass spectrometry, specifically using the matrix
CC   -assisted laser desorption/ionisation (MALDI) or electrospray techniques.
CC   To improve detection in the mass spectrometer, fragments formed in the
CC   instrument have only a single net charge (positive or negative). The
CC   genomic DNA is from e.g. a cell line, biopsy sample, blood, or paraffin-
CC   embedded tissue sample. Oligonucleotides or peptide-nucleic acids that
CC   are complementary to (or identical with) parts of the nuclei acids listed
CC   above may be used as primers for amplification of the nucleic acids or
CC   their complements, and for determining cytosine methylation status and/or
CC   single nucleotide polymorphisms in metastasis-related genes. They can be
CC   used for analysis of diseases associated with methylation of CpG
CC   dinucleotides and to determine (epi)genetic parameters for diagnosis
CC   and/or therapy of disease (or predisposition). The genomic DNA sequences
CC   are useful for diagnosis and therapy of solid tumours and cancer. The
CC   present sequence is the complementary sequence to a bisulphite treated
CC   human gene associated with metastasis. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030148327.
XX
SQ   Sequence 16724 BP; 5336 A; 237 C; 3558 G; 7593 T; 0 U; 0 Other;

  Query Match             2.1%;  Score 42.6;  DB 7;  Length 16724;
  Best Local Similarity   50.2%;  Pred. No. 5.8;
  Matches  135;  Conservative    0;  Mismatches  129;  Indels    5;  Gaps    1;

Qy       1517 CTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCT 1576
             | || |   |||       |||| | ||  |||  | |  | |     |||| | | |  |
Db      16680 CCTTACAAAACAATAAAAATATATTAATTATTATTAAATTCGTAAACAATAAAATTAATT 16621
```

```
Qy       1577 GTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAAT 1636
Db      16620 CCTAAAATCTATAAACTCTAAAAATACAAAAACAATAAAATATAATTCCTAAAAAAAATC 16561

Qy       1637 CTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCAT 1696
Db      16560 TTTCTTAAAAACAAACTAACGCCTAATACAAACAATAAAAAAAAAATAAAAAACAACCAC 16501

Qy       1697 GTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db      16500 TTTAAAAACATTATATATATTTATATATAACA-----AAAAAATAAAAACTTATAAAAAA 16446

Qy       1757 TCTTTATGATGAGAACTATCTTGACTCAA 1785
Db      16445 TAATTACAAACAAAACTAATATTACCCAA 16417
```

```
RESULT 40
ABX39419
ID   ABX39419 standard; cDNA; 416 BP.
XX
AC   ABX39419;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #4584.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-00960352.
XX
PR   12-JAN-1999;   99US-0115707P.
PR   11-JAN-2000; 2000US-00480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and analysis,
PT   cattle breeding, or for genetically improving cattle.
XX
PS   Claim 2; SEQ ID NO 4584; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived from
CC   cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC   nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC   appearing as ABX34836-ABX49947, or complements of them. Also included are
CC   ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC   acid linked to a promoter and a 3' non- translated sequence that
CC   functions in the cell to cause termination of transcription and addition
CC   of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC   (2) determining a level or pattern of a molecule in a bovine cell or
CC   tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC   of the 15112 nucleic acid sequences or its complement or fragment) with a
CC   complementary nucleic acid molecule obtained from the bovine cell or
CC   tissue, where hybridisation between the marker nucleic acid and the
CC   complementary nucleic acid permits the detection of the molecule; and (b)
CC   detecting the level or pattern of the complementary nucleic acid, where
CC   the detection of the complementary nucleic acid is predictive of the
CC   level or pattern of the molecule. The LMFD nucleic acid is used for
CC   determining a level or pattern of a molecule in a bovine cell or tissue.
CC   It is useful for genome mapping, gene identification and analysis, cattle
CC   breeding, preparation of constructs for use in cattle gene expression, or
CC   for genetically improving cattle. The present sequence is one of the
CC   15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC   present sequence was not shown in the specification but was obtained in
CC   electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139
XX
SQ   Sequence 416 BP; 223 A; 1 C; 14 G; 178 T; 0 U; 0 Other;

  Query Match             2.1%;  Score 42.4;  DB 8;  Length 416;
  Best Local Similarity   52.9%;  Pred. No. 1.4;
  Matches   91;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy       1609 TCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCT 1668
Db        134 TTATAAAAAAAAAAAATAAAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGATTTTAA 193

Qy       1669 CAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCA 1728
Db        194 AAATTAAAATATAAAAATTAAAAAATTCATAAAAAAAAAAAATATTTAAAAAAAAAATTT 253

Qy       1729 TTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGA 1780
Db        254 TTAAATTTTTAAAGAAAAATTTTTAAAAATTTTAAAATTATTAAATTATAGA 305
```

```
RESULT 41
AAZ87220
ID   AAZ87220 standard; DNA; 1317 BP.
XX
AC   AAZ87220;
XX
DT   08-MAY-2000  (first entry)
XX
DE   DNA encoding native BoNT serotype A (BoNTA) C-terminal fragment (Hc).
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype A; C-terminal fragment;
KW   Hc; Venezuelan equine encephalitis virus replicon; VEE; botulism;
KW   vaccine; diagnosis; drug screening; ds.
XX
OS   Clostridium botulinum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1317
FT                   /*tag=  a
FT                   /product= "Native botulinum neurotoxin serotype A (BoNTA)
FT                   heavy chain C-terminal fragment"
FT                   /transl_except= (pos:1294. .1314, aa:Gly)
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   P-PSDB; AAY77142.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
```

PS Example 3; Page 52; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. The present sequence represents DNA encoding native
CC BoNTA heavy chain C-terminal fragment (Hc) used in an exemplification of
CC the present invention
XX
SQ Sequence 1317 BP; 560 A; 108 C; 211 G; 438 T; 0 U; 0 Other;

Query Match 2.1%; Score 42.4; DB 3; Length 1317;
Best Local Similarity 49.5%; Pred. No. 2.2;
Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
Db 345 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 404

Qy 1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
Db 405 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 464

Qy 1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
Db 465 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 524

Qy 1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db 525 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 564

RESULT 42
AAV30575
ID AAV30575 standard; DNA; 1546 BP.
XX
AC AAV30575;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C gene in pHisBotA.
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulism; ds.
XX
OS Clostridium botulinum; serotype A.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 108. .1496
FT /*tag= a
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR P-PSDB; AAW68390.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 28; Page 277-278; 428pp; English.
XX
CC This is the DNA sequence of the Clostridium botulinum serotype A toxin C
CC fragment gene contained in plasmid pHisBotA. The encoded toxin A
CC polypeptide (see AAW68390) has a histidine-tagged N-terminal extension.
CC The vector was used to express native (i.e. non-fusion) soluble C
CC fragment in Escherichia coli host cells. The invention relates to
CC recombinant proteins derived from C. botulinum toxins. Methods are
CC provided which allow for the isolation of soluble recombinant proteins
CC free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 1546 BP; 629 A; 163 C; 256 G; 498 T; 0 U; 0 Other;

Query Match 2.1%; Score 42.4; DB 2; Length 1546;
Best Local Similarity 49.5%; Pred. No. 2.4;
Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
Db 524 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 583

Qy 1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
Db 584 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 643

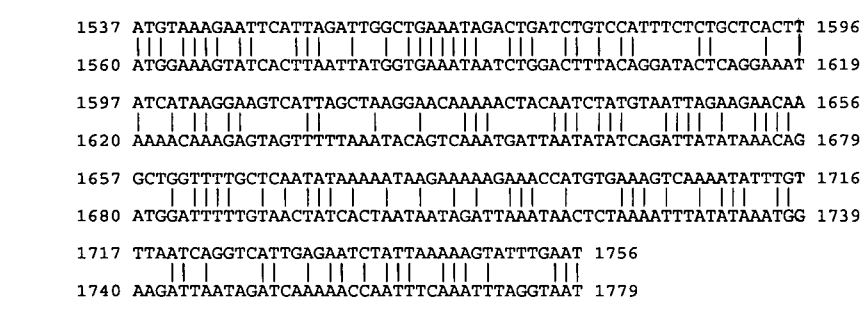
Qy 1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
Db 644 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 703

Qy 1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db 704 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 743

RESULT 43
AAZ87219
ID AAZ87219 standard; DNA; 1917 BP.
XX
AC AAZ87219;
XX
DT 08-MAY-2000 (first entry)
XX
DE DNA encoding native BoNT serotype A (BoNTA) N-terminal fragment (Hn).
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype A; N-terminal fragment;
KW Hn; Venezuelan equine encephalitis virus replicon; VEE; botulism;
KW vaccine; diagnosis; drug screening; ds.
XX
OS Clostridium botulinum.
XX
FH Key Location/Qualifiers
FT CDS 1. .1221
FT /*tag= a
FT /product= "Native botulinum neurotoxin serotype A (BoNTA)
FT heavy chain N-terminal fragment"
FT /note= "No stop codon given in the specification"

XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR P-PSDB; AAY77141.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Example 3; Page 51; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. The present sequence represents DNA encoding native
CC BoNTA heavy chain N-terminal fragment (Hn) used in an exemplification of
CC the present invention
XX
SQ Sequence 1917 BP; 789 A; 185 C; 284 G; 659 T; 0 U; 0 Other;

Query Match 2.1%; Score 42.4; DB 3; Length 1917;
Best Local Similarity 49.5%; Pred. No. 2.6;
Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;



RESULT 44
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;

XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 2.1%; Score 42.4; DB 8; Length 2000;
Best Local Similarity 8.6%; Pred. No. 2.7;
Matches 61; Conservative 328; Mismatches 319; Indels 4; Gaps 2;



```
Qy       1625 AAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGA 1684
Db        558 YMKWKCTWRRCMCYRWGYTMYTTSRSRMMYTGRYKARYTSKRRYMWYKYRKYCWYYYYGM 499

Qy       1685 AAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAA 1744
Db        498 YMKCSYMMRYGYCKACKKCCYAMCWKAAYSGMMMYWYRKYSKWMRMSTKYMWSMWYKKCR 439

Qy       1745 AAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTTT 1804
Db        438 SMKYGAKGCYGCKMWTYCSYGYMKWYTYMGSYKYSRCY---KYMRMYMYKGWMYMMYYSA 382

Qy       1805 TTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTACAGGT 1864
Db        381 YSSMMTWYYYYAKYWKYWYKRRGTMSWYGKSYKKKYCTWWCYMKCMRCYRWRKMMRKKTK 322

Qy       1865 GCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTTA 1924
Db        321 YSKRCYCWRYATCYWCCCYRKRGWYSRRSMMRTAGKWKMRSWSRWCRSYSWYKMYKKMWK 262

Qy       1925 GTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCT 1976
Db        261 KSYYMSYGWARSSGTWSRSAAKRTYKGYSTSRRAKMMRACRMYSACRRYSRT 210
```

```
RESULT 45
AAZ87218
ID   AAZ87218 standard; DNA; 2532 BP.
XX
AC   AAZ87218;
XX
DT   08-MAY-2000  (first entry)
XX
DE   DNA encoding native botulinum neurotoxin serotype A (BoNTA).
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype A;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening; ds.
XX
OS   Clostridium botulinum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2532
FT                   /*tag=  a
FT                   /product= "Native botulinum neurotoxin serotype A (BoNTA)
FT                   heavy chain"
FT                   /transl_except= (pos:2509. .2529, aa:Gly)
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   P-PSDB; AAY77140.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Example 3; Page 49-50; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. The present sequence represents DNA encoding native
CC   BoNTA heavy chain used in an exemplification of the present invention
XX
SQ   Sequence 2532 BP; 1038 A; 241 C; 404 G; 849 T; 0 U; 0 Other;
```

```
  Query Match             2.1%;  Score 42.4;  DB 3;  Length 2532;
  Best Local Similarity   49.5%;  Pred. No. 3;
  Matches  109;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy       1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
Db       1560 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 1619

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
Db       1620 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 1679

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
Db       1680 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 1739

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db       1740 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 1779
```

```
Search completed: January 22, 2006, 04:09:00
Job time : 813.081 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:39:03 ; Search time 245.225 Seconds
(without alignments)
14504.671 Million cell updates/sec

Title: US-09-728-552A-3_COPY_78595_80595
Perfect score: 2001
Sequence: 1 atctttatttgtatgagaaa.....aagtcattctgccagaattc 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	80595	3	US-09-078-294-3	Sequence 3, Appli
	2	1984.8	99.2	18443	3	US-09-078-294-6	Sequence 6, Appli
	3	1895.2	94.7	80246	3	US-09-078-294-4	Sequence 4, Appli
	4	97.6	4.9	300	3	US-09-078-294-22	Sequence 22, Appl
	5	60.8	3.0	7218	2	US-08-232-463-14	Sequence 14, Appl
c	6	58.4	2.9	7218	2	US-08-232-463-14	Sequence 14, Appl
	7	49.4	2.5	431	3	US-09-513-999C-19284	Sequence 19284, A
c	8	47	2.3	1141	3	US-09-806-708B-22	Sequence 22, Appl
	9	46.8	2.3	601	3	US-09-949-016-203912	Sequence 203912,
c	10	46.8	2.3	3338	3	US-09-949-016-5761	Sequence 5761, Ap
c	11	46.8	2.3	38119	3	US-09-949-016-17503	Sequence 17503, A
	12	46	2.3	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	13	43.8	2.2	35380	3	US-09-949-016-11994	Sequence 11994, A
	14	42.4	2.1	1327	3	US-09-350-756-9	Sequence 9, Appli
	15	42.4	2.1	3891	2	US-08-480-604A-27	Sequence 27, Appl
	16	42.4	2.1	3891	2	US-08-405-496A-27	Sequence 27, Appl
	17	42.4	2.1	3891	3	US-08-915-136-27	Sequence 27, Appl
	18	42.4	2.1	3891	3	US-09-084-517-27	Sequence 27, Appl
	19	42.4	2.1	4835	3	US-09-288-326A-10	Sequence 10, Appl
	20	42.4	2.1	4835	3	US-09-548-409B-10	Sequence 10, Appl
	21	42.2	2.1	601	3	US-09-949-016-203911	Sequence 203911,
	22	42	2.1	601	3	US-09-949-016-203451	Sequence 203451,
	23	42	2.1	601	3	US-09-949-016-203452	Sequence 203452,
	24	42	2.1	113060	3	US-09-949-016-14773	Sequence 14773, A
	25	42	2.1	113060	3	US-09-949-016-14774	Sequence 14774, A
c	26	41.6	2.1	832	3	US-09-621-976-2813	Sequence 2813, Ap
c	27	41.2	2.1	3312	2	US-08-049-473-1	Sequence 1, Appli
c	28	41.2	2.1	3312	2	US-08-312-648-1	Sequence 1, Appli
c	29	41.2	2.1	3312	6	PCT-US94-04190-1	Sequence 1, Appli
c	30	41.2	2.1	113060	3	US-09-949-016-14773	Sequence 14773, A
c	31	41.2	2.1	113060	3	US-09-949-016-14774	Sequence 14774, A
c	32	41	2.0	312470	3	US-09-949-016-14043	Sequence 14043, A
c	33	40.8	2.0	92155	3	US-09-949-016-17484	Sequence 17484, A
c	34	40.6	2.0	601	3	US-09-949-016-86304	Sequence 86304, A
c	35	40.6	2.0	601	3	US-09-949-016-86305	Sequence 86305, A
	36	40.6	2.0	422592	3	US-09-949-016-14182	Sequence 14182, A
c	37	40.6	2.0	640681	3	US-09-790-988-1	Sequence 1, Appli
c	38	40.4	2.0	252	3	US-09-248-796A-10353	Sequence 10353, A
c	39	39.6	2.0	476	3	US-09-573-080A-118	Sequence 118, App
	40	39.6	2.0	56374	3	US-09-949-002-645	Sequence 645, App
	41	39.6	2.0	56375	3	US-09-949-002-774	Sequence 774, App
	42	39.4	2.0	1477	3	US-08-096-181A-7	Sequence 7, Appli
	43	39.4	2.0	1477	6	PCT-US94-08326-7	Sequence 7, Appli
	44	39.2	2.0	1987	3	US-09-350-756-8	Sequence 8, Appli
	45	39.2	2.0	2452	3	US-09-350-756-7	Sequence 7, Appli
	46	39.2	2.0	2613	3	US-09-255-829-7	Sequence 7, Appli
	47	39.2	2.0	2616	3	US-09-255-829-1	Sequence 1, Appli
	48	39.2	2.0	2616	3	US-09-255-829-25	Sequence 25, Appl
	49	39.2	2.0	2622	3	US-09-255-829-5	Sequence 5, Appli
	50	39.2	2.0	2628	3	US-09-255-829-9	Sequence 9, Appli
	51	39.2	2.0	2637	3	US-09-255-829-11	Sequence 11, Appl
	52	39.2	2.0	2685	3	US-09-255-829-3	Sequence 3, Appli
	53	39.2	2.0	2724	3	US-09-255-829-15	Sequence 15, Appl
	54	39.2	2.0	2862	3	US-09-255-829-13	Sequence 13, Appl
	55	39.2	2.0	3042	3	US-09-255-829-17	Sequence 17, Appl
	56	39.2	2.0	98701	3	US-09-949-016-15898	Sequence 15898, A
	57	39.2	2.0	98701	3	US-09-949-016-15899	Sequence 15899, A
	58	39.2	2.0	129380	3	US-09-949-016-12544	Sequence 12544, A
	59	39.2	2.0	136917	3	US-09-949-016-16369	Sequence 16369, A
	60	39	1.9	8229	3	US-09-949-016-11983	Sequence 11983, A
	61	39	1.9	8376	3	US-09-949-016-15561	Sequence 15561, A
c	62	39	1.9	19124	2	US-08-487-826B-13	Sequence 13, Appl
	63	39	1.9	60276	3	US-09-949-016-15004	Sequence 15004, A
	64	39	1.9	60338	3	US-09-949-016-15694	Sequence 15694, A
c	65	38.6	1.9	601	3	US-09-949-016-78775	Sequence 78775, A
c	66	38.6	1.9	4115	3	US-09-302-620B-85	Sequence 85, Appl
c	67	38.6	1.9	4115	3	US-09-912-161-7	Sequence 7, Appli
	68	38.6	1.9	10327	3	US-09-949-002-804	Sequence 804, App
	69	38.6	1.9	16677	3	US-09-949-002-625	Sequence 625, App
c	70	38.6	1.9	18773	3	US-09-949-016-14164	Sequence 14164, A
	71	38.6	1.9	331814	3	US-09-949-016-12008	Sequence 12008, A
	72	38.6	1.9	331814	3	US-09-949-016-17056	Sequence 17056, A
c	73	38.6	1.9	784019	3	US-09-949-016-14033	Sequence 14033, A
c	74	38.6	1.9	828152	3	US-09-949-016-12777	Sequence 12777, A
	75	38.4	1.9	52789	3	US-09-949-016-12130	Sequence 12130, A
	76	38.4	1.9	52790	3	US-09-949-016-16641	Sequence 16641, A
c	77	38.4	1.9	103750	3	US-09-949-016-13319	Sequence 13319, A
c	78	38.4	1.9	105050	3	US-09-949-016-15953	Sequence 15953, A
	79	38.4	1.9	451924	3	US-09-949-016-12896	Sequence 12896, A
	80	38.4	1.9	451925	3	US-09-949-016-17305	Sequence 17305, A
	81	38.2	1.9	885	3	US-08-952-089A-2	Sequence 2, Appli
	82	38.2	1.9	885	3	US-09-690-885-2	Sequence 2, Appli
	83	38.2	1.9	1588	3	US-08-952-089A-26	Sequence 26, Appl
	84	38.2	1.9	1588	3	US-09-690-885-26	Sequence 26, Appl
	85	38.2	1.9	4146	3	US-08-952-089A-29	Sequence 29, Appl
	86	38.2	1.9	4146	3	US-09-690-885-29	Sequence 29, Appl
c	87	38	1.9	4434	3	US-09-830-230A-416	Sequence 416, App
c	88	38	1.9	4485	3	US-09-830-230A-415	Sequence 415, App
	89	38	1.9	32099	3	US-09-949-016-16562	Sequence 16562, A
	90	38	1.9	640681	3	US-09-790-988-1	Sequence 1, Appli
c	91	37.8	1.9	1055	3	US-09-806-708B-23	Sequence 23, Appl
	92	37.8	1.9	1944	3	US-09-710-279-567	Sequence 567, App
c	93	37.8	1.9	3353	3	US-09-710-279-3481	Sequence 3481, Ap
	94	37.8	1.9	3558	3	US-09-134-001C-2439	Sequence 2439, Ap
c	95	37.8	1.9	47284	3	US-09-949-016-17029	Sequence 17029, A
c	96	37.6	1.9	32392	3	US-09-662-254B-27	Sequence 27, Appl
	97	37.6	1.9	68283	3	US-09-949-016-12261	Sequence 12261, A

c	98	37.6	1.9	83851	3	US-09-949-016-13847	Sequence 13847, A
	99	37.6	1.9	144596	3	US-09-949-016-11749	Sequence 11749, A
	100	37.6	1.9	144596	3	US-09-949-016-13035	Sequence 13035, A

ALIGNMENTS

RESULT 1
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match 100.0%; Score 2001; DB 3; Length 80595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78595 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 78654

Qy         61 TCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78655 TCAGGACCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAA 78714

Qy        121 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78715 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAAC 78774

Qy        181 AAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78775 AAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCT 78834

Qy        241 GTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78835 GTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGA 78894

Qy        301 TGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78895 TGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTTT 78954

Qy        361 TTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78955 TTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAAC 79014

Qy        421 ATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79015 ATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTTA 79074

Qy        481 CTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79075 CTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAAG 79134

Qy        541 CTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79135 CTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCTT 79194

Qy        601 TTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79195 TTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTTT 79254

Qy        661 TTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79255 TTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCA 79314

Qy        721 GGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79315 GGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAG 79374

Qy        781 CATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79375 CATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTG 79434

Qy        841 ACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79435 ACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACAC 79494

Qy        901 GAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79495 GAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGCT 79554

Qy        961 GTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79555 GTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCAC 79614

Qy       1021 TAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79615 TAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGAC 79674

Qy       1081 CTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79675 CTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTC 79734

Qy       1141 CATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79735 CATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGACA 79794

Qy       1201 GTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79795 GTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAAT 79854

Qy       1261 GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79855 GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGCA 79914

Qy       1321 TAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAAT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79915 TAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAAT 79974

Qy       1381 ACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79975 ACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCTT 80034

Qy       1441 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80035 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 80094

Qy       1501 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80095 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 80154

Qy       1561 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80155 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 80214

Qy       1621 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80215 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 80274

Qy       1681 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      80275 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 80334

Qy       1741 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80335 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 80394

Qy       1801 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTAC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80395 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTAC 80454

Qy       1861 AGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80455 AGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACT 80514

Qy       1921 TTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80515 TTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGT 80574

Qy       1981 GAAGTCATTCTGCCAGAATTC 2001
              |||||||||||||||||||||
Db      80575 GAAGTCATTCTGCCAGAATTC 80595


RESULT 2
US-09-078-294-6
; Sequence 6, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 18443
;   TYPE: DNA
;   ORGANISM: BAC-F2 contig 2
US-09-078-294-6

  Query Match             99.2%;  Score 1984.8;  DB 3;  Length 18443;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1997;  Conservative    1;  Mismatches    3;  Indels    1;  Gaps    1;

Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db       4292 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATTGT 4351

Qy         61 TCAGGA-CCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 119
              |||| | || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4352 TCAGAACCCTTGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 4411

Qy        120 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4412 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 4471

Qy        180 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4472 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 4531

Qy        240 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4532 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 4591

Qy        300 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4592 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 4651

Qy        360 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       4652 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 4711

Qy        420 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4712 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 4771

Qy        480 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4772 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 4831

Qy        540 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4832 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 4891

Qy        600 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4892 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 4951

Qy        660 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4952 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 5011

Qy        720 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5012 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 5071

Qy        780 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5072 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 5131

Qy        840 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5132 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 5191

Qy        900 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5192 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 5251

Qy        960 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5252 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 5311

Qy       1020 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5312 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 5371

Qy       1080 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5372 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 5431

Qy       1140 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5432 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 5491

Qy       1200 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5492 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 5551

Qy       1260 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5552 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 5611

Qy       1320 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5612 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 5671

Qy       1380 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5672 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 5731

Qy       1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5732 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 5791
```

```
Qy       1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5792 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 5851

Qy       1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5852 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 5911

Qy       1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5912 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 5971

Qy       1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 1739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5972 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 6031

Qy       1740 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 1799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6032 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 6091

Qy       1800 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 1859
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6092 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 6151

Qy       1860 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGAC 1919
             ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db       6152 CAGGTGCCAAGTTAAGTGCCCAAGCTTGMTCTTAAAAGCATACTGGATTTTGTTTTAGAC 6211

Qy       1920 TTTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGG 1979
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6212 TTTTAGTGAACTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGG 6271

Qy       1980 TGAAGTCATTCTGCCAGAATTC 2001
             |||||||||||||||||||||||
Db       6272 TGAAGTCATTCTGCCAGAATTC 6293
```

RESULT 3
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 94.7%; Score 1895.2; DB 3; Length 80246;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1908; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy          1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db      78335 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATTGT 78394

Qy         61 TCAGGA-CCATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 119
              |||| | || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78395 TCAGAACCCTTGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGA 78454

Qy        120 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78455 AAATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAA 78514

Qy        180 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78515 CAAGAGAATTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCC 78574

Qy        240 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78575 TGTCCCTGAGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGG 78634

Qy        300 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78635 ATGAAGGTAGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGGTTTT 78694

Qy        360 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78695 TTTTGATGTCTTCCTTAATTTCACAAATGCTTCCAACAAAGTAGCACACAGGAAAAAGAA 78754

Qy        420 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78755 CATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGCTTTTCACAGCTT 78814

Qy        480 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78815 ACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAGGGTTTCTACTAA 78874

Qy        540 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78875 GCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAGTGATCTTTCTCT 78934

Qy        600 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78935 TTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCTTTGCTTTGGGTT 78994

Qy        660 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78995 TTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCC 79054

Qy        720 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79055 AGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAA 79114

Qy        780 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79115 GCATTCTTGGCAAAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTT 79174

Qy        840 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79175 GACATGTTGTAAGATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACA 79234

Qy        900 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79235 CGAACTAAGCTGCTCCTGGAATGCAGGGTGGCCTCCTCAATACAGGATGTTCTAGAGAGC 79294

Qy        960 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79295 TGTATTTTGGGCACTTAACTATTCTCCACTACTTAGGGCACAGCACTGAAATTAACACCA 79354

Qy       1020 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79355 CTAAGTTTGTCATGTCCATGTAGTTAGTCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGA 79414

Qy       1080 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79415 CCTCTTATGTGTGTCCAGCCTTTCTTCCTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCT 79474

Qy       1140 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79475 CCATAGGAACATCAGTCCTGAATCCTCAGACCACCATCTGGAGTAGTAAGTGCTCCTGAC 79534

Qy       1200 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79535 AGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAAGCGTGTGACACACCGTGAGAGAGAAA 79594
```

```
Qy       1260 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 1319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79595 TGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAGCCCCCTAATTTTACTGC 79654

Qy       1320 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 1379
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79655 ATAATTATTTTGAATTCACTGATAATTTCTACAATTTTCCCATAAGTCATCTACACACAA 79714

Qy       1380 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 1439
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79715 TACCCTCTCATGCAACACTTGGCTTTGCTAATACATATCTATTATGAGAGCTGTGCTTCT 79774

Qy       1440 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 1499
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79775 TAAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAG 79834

Qy       1500 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 1559
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79835 CAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGG 79894

Qy       1560 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 1619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79895 CTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAA 79954

Qy       1620 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 1679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79955 GGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAA 80014

Qy       1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 1739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80015 TAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTA 80074

Qy       1740 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 1799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80075 TTAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGA 80134

Qy       1800 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 1859
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80135 GCTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAAGTCTTATATGGTA 80194

Qy       1860 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 1911
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80195 CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTG 80246
```

RESULT 4
US-09-078-294-22
; Sequence 22, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 300
; TYPE: DNA
; ORGANISM: BAC-F2 contig 47 fragment 2
US-09-078-294-22

Query Match 4.9%; Score 97.6; DB 3; Length 300;
Best Local Similarity 66.7%; Pred. No. 1.1e-17;
Matches 168; Conservative 18; Mismatches 62; Indels 4; Gaps 4;

```
Qy       1753 GAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAGCTTTTTGGCCTG 1812
             ||||||| | ::|| |:||||||||||:||||| | :||   |             |||||||
Db          1 GAATTCTCTGWKATTAKAACTATCTTGMCTCAAATTSACTTGGTGAGCTAACCTGGCCTG 60
```

```
Qy       1813 TGGTCCCTACG-TAGAAAGGAGGCTTTGTCATAAAG-TCTTATATGGTACAGGTGCCAAG 1870
             |||||||| | |  || ||||||||||||||| || || |:| |||||| :|||| ||
Db         61 TGGTCCCTTGGCTTTAATGGAGGCTTTGTCATATAGATCATMTGTGGTACTKGTGCCTAG 120

Qy       1871 TT-AAGTGCCCAAGCTTGCTCTTAAAAGCATACTGGATTTTGTTTTAGACTTTTAGTGAA 1929
             ||  |||||||  .||||||:||  : || ||||:||||:|   || || | :| : ||
Db        121 TTGTAGTGCCCTGCCTTGCTSTTCTWGGCTTACTKGATTTWGGGGTATACATCWATKTAA 180

Qy       1930 CTGAAGGGAATAAACAAATCCCTCTGGGAGAACTTCTCCTCCATCCTTGGTGAAGTCATT 1989
             :|:|| ||  |   |   |||| ::|||||| |||||||||  ||| || ||| || ||
Db        181 YTSAAAGGTCT-TTCTCCTCCCGYYGGGAGAATTTCTCCTCCTCCCTCGGAGAACTCTTT 239

Qy       1990 CTGCCAGAATTC 2001
             ||:||  |||||
Db        240 CTSCCGAAATTC 251
```

RESULT 5
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 3.0%; Score 60.8; DB 2; Length 7218;
Best Local Similarity 2.9%; Pred. No. 2e-06;
Matches 11; Conservative 226; Mismatches 143; Indels 0; Gaps 0;

```
Qy        408 CAGGAAAAAGAACATAGGGACTCTACTGGTGGGTGCTTTTATCTTAAGCCTTGTACTTGC 467
              ||||   | | |  | | || ::: ::  :   : ::::: ::::   :::: : ::: :
Db       1046 CAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1105
```

```
Qy        468 TTTTCACAGCTTACTCACTGCTTGTACCTGAGGCCATATGCCCTGTAAAAGCTTCTGCAG 527
              ::::: :  ::: ::: :: ::: : :::    :: : : :::: :     ::::: :
Db       1106 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1165

Qy        528 GGTTTCTACTAAGCTGGGTTCCTTATATGGCTCTCTCCCATTTCTGTTGCCTCACTCTAG 587
              ::::: ::   ::   :::::: : :  ::::::::: ::::: :: :::: ::::
Db       1166 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1225

Qy        588 TGATCTTTCTCTTTTCCTCACCTCTGGGACTGGTGGCTGTTTGTATGGACTGCCTTAGCT 647
              :  :::::::::::::::: :::::    ::  :  :: ::: : :   :: :::::  ::
Db       1226 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1285

Qy        648 TTGCTTTGGGTTTTTTCCTGGGGACAATGTCTTCAGATTATCCTAGACCAAATAAACTAC 707
              :: ::::   :::::::::      :  : :::::   :: ::::   ::   :   :: :
Db       1286 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1345

Qy        708 AGCCACTGGGCCAGGCTCTTCCTCCTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTT 767
              :: ::   ::   :::::::::::::  ::   :: : :::::    :::::: ::::
Db       1346 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1405

Qy        768 AGTTTAGGTCAAGCATTCTT 787
              :::   ::   : :::::
Db       1406 YYYYYYYYYYYYYYYYYYY 1425


RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             2.9%;  Score 58.4;  DB 2;  Length 7218;
  Best Local Similarity   1.4%;  Pred. No. 9.5e-06;
  Matches    5;  Conservative  216;  Mismatches  127;  Indels    0;  Gaps    0;

Qy          8 TTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCAGGAC 67
              || ||| :::::::::::::::::: :::::: : :  :: : : : : ::   ::::
Db       1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

Qy         68 CATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATAT 127
               : :::  : ::: :: :::::   ::   :::  :: : :::  :::::::::: : :
Db       1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

Qy        128 TGGCTCATGTAACCTTCTCACAGAGAGGGCAGGATGGAAGGGGCTTTGGGAACAAGAGAA 187
               ::   : : ::       : :::::::: :::: :::::::::    ::::: :::::::
Db       1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

Qy        188 TTGTTCTCAAATTCTAGGAATACTAGGATTAGTCCAGGATGGGTCACCTTCCTGTCCCTG 247
               :     :::    ::::: :  :::  ::   :::: :::  :       :     :
Db       1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

Qy        248 AGGTGGTGGTAGCGATGGTAGAGTCTTATGGGAGGAAAGAGTGCATGTTAGGATGAAGGT 307
              ::: :: :: :: :: :: :::    : :::::::::::: : : :  :::: :::::
Db       1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

Qy        308 AGGGCTAAGCAAACAAGGGCAAGGGCCACTATATCATGCTAAAAATGG 355
              ::::  ::: ::: ::::: :::::  :  : :  : :  ::::: ::
Db       1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1094


RESULT 7
US-09-513-999C-19284
; Sequence 19284, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 19284
;   LENGTH: 431
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 425
;   OTHER INFORMATION: r=a or g
US-09-513-999C-19284

  Query Match             2.5%;  Score 49.4;  DB 3;  Length 431;
  Best Local Similarity   71.4%;  Pred. No. 0.00081;
  Matches   65;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         70 TGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATATTG 129
              | | ||| |||||||||||||| ||| |||||  | | || | ||||| || ||||||| |
Db        231 TTGGTTGTAAGTGACAGAAACCCAACTCAAATTAGCTTAGGCCAAAAAGAGAATATATGG 290

Qy        130 GCTCATGTAACCTTCTCACAGAGAGGGCAGG 160
              | ||||| ||||    |  ||| | || | |
Db        291 GTTCATGGAACCCAAACTAAGAAAAGGGACG 321


RESULT 8
US-09-806-708B-22/c
```

```
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             2.3%;  Score 47;  DB 3;  Length 1141;
  Best Local Similarity   12.8%;  Pred. No. 0.0066;
  Matches   75;  Conservative  205;  Mismatches  303;  Indels    3;  Gaps    2;

Qy        807 TAACAATAGACATTCTAGCAATTGATTCTTTTTGACATGTTGTAAGATCTATTCACATTT 866
              :|       |::| |: : |   :: |:: ::|   : :    :: :|| :|::
Db        926 KAKNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRTRHHTRTCRRTKYNNNNNN 867

Qy        867 TGTAATTAAAGCATTCCCCTATGGAAACCAACACGAACTAAGCTGCTCCTGGAATGCAGG 926
               :|: ::::|  ::     |:|          ||:   |:: ::::  |
Db        866 ARTVYWYHHAARRWMNAWWTRTNNNNNNNNNNACRNTRTWWABWKHSWCNNNNNNNNNN 807

Qy        927 GTGGCCTCCTCAATACAGGATGTTCTAGAGAGCTGTATTTTGGGCACTTAACTATTCTCC 986
                 |:: |  |: | :        : |:      :   |    :  :     |::: :
Db        806 NNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMHAAVTTTHTDWCYKTWMNTWYWDM 747

Qy        987 ACTACTTAGGGCACAGCACTGAAATTAACACCACTAAGTTTGTCATGTCCATGTAGTTAG 1046
              : |::||  : :     :         :||       :  :: ::|: |  :  ::  |:
Db        746 MTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNNNM--WKAYYAHATNNWGCWWNNTDA 689

Qy       1047 TCTCAGGCAGTGCAGCCTCAGGAGTGGAACTGACCTCTTATGTGTGTCCAGCCTTTCTTC 1106
               |    ::: : :   : :    |:::::|||                 :||| |
Db        688 RRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNNNNNNNNSCCTCTRMMT 629

Qy       1107 CTTCAGAAGTCAGCTGTGTTTTCTGCTGACTCTCCATAGGAACATCAGTCCTGAATCCTC 1166
              : : :: :|  :: :::  :|||| :::: :::: ::|  ::|:::  |       ::: :
Db        628 MRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCRDVTYTRNNTYCKSYAH 569

Qy       1167 AGACCACCATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGGAT 1226
                ::    |  ::: : : :    : |    : :: :: :::|||  : ::    |: ::
Db        568 SYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNNTDTRY 509

Qy       1227 CTCCAAAGCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAAT 1286
              ::   ::  :| | :::::|  ::: :|  ::| :: :||  : | ::::|:: |   ::
Db        508 YWWWKRWARBT-TTVYDSMCNAKSMWRGNNWRAMKMWWAANNDAGAMDHWTYWMGNNTMW 450

Qy       1287 CTTGCTTTTTCACAAGCCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAATT 1346
              :    :     :|:: :||          :::  ::::|: : :: |     :: ::::
Db        449 MRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWKYMWKAACNNNNBKAMYMRVA 390

Qy       1347 TCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCATGC 1392
              ::  :::|| |  :::  ::::: : ::: ::: :       : :|
Db        389 WMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNNNNWRBC 344


RESULT 9
US-09-949-016-203912
; Sequence 203912, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 203912
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-203912

  Query Match             2.3%;  Score 46.8;  DB 3;  Length 601;
  Best Local Similarity   49.7%;  Pred. No. 0.0054;
  Matches  172;  Conservative    0;  Mismatches  172;  Indels    2;  Gaps    2;

Qy       1446 TAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGT 1505
              | |||| ||| | | |   ||       | | ||  | | ||||   |  ||||  |   |
Db        116 TTAATGCTTTGTTTTCTAAAACTTGTATTGTTTTTAAACAAAAATAGATTTGAAAAGAAT 175

Qy       1506 GATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAA 1565
              ||  || ||   |  | | | | | ||| ||| |     |  | ||| |  |   || | |
Db        176 GAACCACAACAGTGCTACCAAAAAGTCACATCTTGTTACCAGACATCA-CTGTCCTTACA 234

Qy       1566 TAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACA 1625
              | | | ||    |  | | | |     ||||    | ||        |   || |
Db        235 ACAATTCAACTAATAAGATTTAAGATTTAACTTCATCCTAATGTATCAGAACATTAAAAA 294

Qy       1626 AAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAA 1685
              ||||  |   || |||||  ||    || ||   |   || |  |||  |   |  ||
Db        295 AAAAAAAGTCTCAATGTACCTATTTAAATAAATCGTGCTTCCATAATAATATTCTTTAAA 354

Qy       1686 AAAGAAACCATGTGAA-AGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAA 1744
              || ||||    || || ||   | || ||| |||| |||   || |||  |  || | |
Db        355 AATGAAAAACAGTAAAGAGCACACATTTTTATTTACTCACAACACTGAATAATTAATGTA 414

Qy       1745 AAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGA 1790
              || | ||||||||| ||| |      |   || ||  ||| || | |
Db        415 AACTTTTTGAATTTTTTTTTTCTTTAGACATTTTTCCTCTAGAGTA 460


RESULT 10
US-09-949-016-5761/c
; Sequence 5761, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5761
;   LENGTH: 3338
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-5761
```

```
  Query Match             2.3%;  Score 46.8;  DB 3;  Length 3338;
  Best Local Similarity   49.7%;  Pred. No. 0.013;
  Matches  172;  Conservative    0;  Mismatches  172;  Indels    2;  Gaps    2;

Qy       1446 TAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGT 1505
             | |||| ||| | | |   ||       | | ||  | | ||||   |  ||||  |   |
Db       3142 TTAATGCTTTGTTTTCTAAAACTTGTATTGTTTTTAAACAAAAATAGATTTGAAAAGAAT 3083

Qy       1506 GATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAA 1565
             ||  || ||   |  |  | | | ||| |||  |     |  ||| |  |   || | |
Db       3082 GAACCACAACAGTGCTACCAAAAAGTCACATCTTGTTACCAGACATCA-CTGTCCTTACA 3024

Qy       1566 TAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACA 1625
                | | | ||    |  | |  | |       ||||    | ||        |   || |
Db       3023 ACAATTCAACTAATAAGATTTAAGATTTAACTTCATCCTAATGTATCAGAACATTAAAAA 2964

Qy       1626 AAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAA 1685
              ||||  |   || |||||  ||    || ||   |   || |  ||||  |   |   ||
Db       2963 AAAAAAAGTCTCAATGTACCTATTTAAATAAATCGTGCTTCCATAATAATATTCTTTAAA 2904

Qy       1686 AAAGAAACCATGTGAA-AGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAA 1744
              || ||||    || || ||   | || ||| |||| |||   || |||  |  || |  |
Db       2903 AATGAAAAACAGTAAAGAGCACACATTTTTATTTACTCACAACACTGAATAATTAATGTA 2844

Qy       1745 AAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGA 1790
              || | |||||||| ||| |       |   || ||  ||| || | |
Db       2843 AACTTTTGAATTTTTTTTTTCTTTAGACATTTTCCTCTAGAGTA 2798


RESULT 11
US-09-949-016-17503/c
; Sequence 17503, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17503
;   LENGTH: 38119
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17503

  Query Match             2.3%;  Score 46.8;  DB 3;  Length 38119;
  Best Local Similarity   49.7%;  Pred. No. 0.047;
  Matches  172;  Conservative    0;  Mismatches  172;  Indels    2;  Gaps    2;

Qy       1446 TAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGT 1505
             | |||| ||| | | |   ||       | | ||  | | ||||   |  ||||  |   |
Db      35922 TTAATGCTTTGTTTTCTAAAACTTGTATTGTTTTTAAACAAAAATAGATTTGAAAAGAAT 35863

Qy       1506 GATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAA 1565
             ||  || ||   |  |  | | | ||| |||  |     |  ||| |  |   || | |
Db      35862 GAACCACAACAGTGCTACCAAAAAGTCACATCTTGTTACCAGACATCA-CTGTCCTTACA 35804

Qy       1566 TAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACA 1625
                | | | ||    |  | |  | |       ||||    | ||        |   || |
Db      35803 ACAATTCAACTAATAAGATTTAAGATTTAACTTCATCCTAATGTATCAGAACATTAAAAA 35744

Qy       1626 AAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAA 1685
              ||||  |   || |||||  ||    || ||   |   || |  ||||  |   |   ||
Db      35743 AAAAAAAGTCTCAATGTACCTATTTAAATAAATCGTGCTTCCATAATAATATTCTTTAAA 35684

Qy       1686 AAAGAAACCATGTGAA-AGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAA 1744
              || ||||    || || ||   | || ||| |||| |||   || |||  |  || |  |
Db      35683 AATGAAAAACAGTAAAGAGCACACATTTTTATTTACTCACAACACTGAATAATTAATGTA 35624

Qy       1745 AAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGA 1790
              || | |||||||| ||| |       |   || ||  ||| || | |
Db      35623 AACTTTTTGAATTTTTTTTTTCTTTAGACATTTTTCCTCTAGAGTA 35578


RESULT 12
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             2.3%;  Score 46;  DB 3;  Length 1141;
  Best Local Similarity   13.1%;  Pred. No. 0.013;
  Matches   77;  Conservative  223;  Mismatches  278;  Indels   11;  Gaps    2;

Qy       1174 CATCTGGAGTAGTAAGTGCTCCTGACAGTCCTAGAAGTTGTCTACCGCTGGATCTCCAAA 1233
              :: :: :::: :::: :|::  : :: :  ::: ::::  : ::   :  :  :  :::|
Db        195 MVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVA 254

Qy       1234 GCGTGTGACACACCGTGAGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTT 1293
              :    :::     :  | ::: :       | | |:: ::: :  |::    | ::: ||
Db        255 WBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKSKTT 314

Qy       1294 TTTCACAAGCCCCCTAATTTTACTGCATAATTATTTTGAATTCACTGATAATTTCTACAA 1353
              :   |:|  |   :| ::::: : :| :  ::       :|:   :::|:::::::|:
Db        315 KVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNWTYKKARHBARWDWVWHSAW 374

Qy       1354 TTTTCCCATAAGTCATCT----------ACACACAATACCCTCTCATGCAACACTTGGCT 1403
              ::::   |:  :: :|::           :: :::| : :::::::        || |
Db        375 KKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRT 434

Qy       1404 TTGCTAATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCAC 1463
              ::| |  : ::| : :     :::|  ::|  :  ||:::::| :   ::: :|  :::
Db        435 YYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGKSHR 494

Qy       1464 TAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCT 1523
              :|| ::| ::   : ::|:|    ::   :| |  :: : :: :  :| :  :: : :
Db        495 BAAAVYT-WYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWWNNYAAWYTKSSWNYTS 553

Qy       1524 CCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATT 1583
               :  ::    ::: :|: :|:|   :|       ::  |::: :    :       ::: :
Db        554 RYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHTWBHBRAGAAHYWMB 613

Qy       1584 TCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTA 1643
              ::: :|:|:  : :  : ||   :                 |  |: : :|| :
Db        614 MMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNNATCARDDYYAASRWYAMAN 673

Qy       1644 ATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAG 1703
              |::    :||  | :|    ::||:  ||:|:: :::       |  :        : |
Db        674 AKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWKNNNNNNAGTWKNNNNNNAKNASA 733
```

```
Qy       1704 TCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTT 1752
Db        734 AKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAYTKYTT 782


RESULT 13
US-09-949-016-11994/c
; Sequence 11994, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11994
;   LENGTH: 35380
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(35380)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11994

  Query Match             2.2%;  Score 43.8;  DB 3;  Length 35380;
  Best Local Similarity   47.9%;  Pred. No. 0.33;
  Matches  126;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy       1531 TCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGC 1590
Db       8347 TCTCATTTAAAATTAAAGCTTTTAAATCGCAGAGCTAAAAAGAACTTACTATTTCATTTC 8288

Qy       1591 TCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAA 1650
Db       8287 ATAGATAAAGAAAGGCAGCCTAGAGAGATTAAAGGGTTTCCCTAAAATCAGTTAAAAGAT 8228

Qy       1651 GAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAAT 1710
Db       8227 TCCATTTCTGGAAAATTTCAAAATAAAACTGAAACATACACAACATCCTCATATAGAGAG 8168

Qy       1711 ATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGA 1770
Db       8167 GATTGGTTACACAGATGTGGGGCATTCAAATACAATTTCCAACAAACTTTTGCATTTGTA 8108

Qy       1771 ACTATCTTGACTCAAGTGGACAG 1793
Db       8107 AAGTGCTTTACAACAGACAAAAG 8085


RESULT 14
US-09-350-756-9
; Sequence 9, Application US/09350756
; Patent No. 6495143
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Reseach Institute for Infectious  Diseases
;  APPLICANT: John S. Lee
;  APPLICANT: Peter Pushko
;  APPLICANT: Michael D. Parker
;  APPLICANT: Jonathan F. Smith
;  APPLICANT: Mark T. Dertzbaugh
;  APPLICANT: Leonard Smith
;  TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
;  FILE REFERENCE: 003/124/SAP RIID 98-21
;  CURRENT APPLICATION NUMBER: US/09/350,756
;  CURRENT FILING DATE: 1999-07-09
;  EARLIER APPLICATION NUMBER: US 60/092,416
;  EARLIER FILING DATE: 1998-07-10
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 9
;   LENGTH: 1327
;   TYPE: DNA
;   ORGANISM: Clostridium botulinum
;   FEATURE:
US-09-350-756-9

  Query Match             2.1%;  Score 42.4;  DB 3;  Length 1327;
  Best Local Similarity   49.5%;  Pred. No. 0.15;
  Matches  109;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy       1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
Db        345 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 404

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
Db        405 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 464

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
Db        465 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 524

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db        525 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 564


RESULT 15
US-08-480-604A-27
; Sequence 27, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3891 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3888
US-08-480-604A-27

  Query Match             2.1%;  Score 42.4;  DB 2;  Length 3891;
  Best Local Similarity   49.5%;  Pred. No. 0.26;
  Matches  109;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy       1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
             ||| |||| ||   ||| |  | ||||||  |||  || | ||     ||    |  |
Db       2919 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 2978

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
             |  | || ||     ||    |   |  |||     ||| |||  |||| |  ||||
Db       2979 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 3038

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
                | ||||  | | ||| | |  |  | | |  |||  |    ||| |  | |||  ||
Db       3039 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 3098

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
                || |   ||  | ||| | |||  ||| |    |||
Db       3099 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 3138


RESULT 16
US-08-405-496A-27
; Sequence 27, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3891 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3888
US-08-405-496A-27

  Query Match             2.1%;  Score 42.4;  DB 2;  Length 3891;
  Best Local Similarity   49.5%;  Pred. No. 0.26;
  Matches  109;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy       1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
             ||| |||| ||   ||| |  | ||||||  |||  || | ||     ||    |  |
Db       2919 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 2978

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
             |  | || ||     ||    |   |  |||     ||| |||  |||| |  ||||
Db       2979 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 3038

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
                | ||||  | | ||| | |  |  | | |  |||  |    ||| |  | |||  ||
Db       3039 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 3098

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
                || |   ||  | ||| | |||  ||| |    |||
Db       3099 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 3138


RESULT 17
US-08-915-136-27
; Sequence 27, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3891 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3888
US-08-915-136-27

  Query Match             2.1%;  Score 42.4;  DB 3;  Length 3891;
  Best Local Similarity   49.5%;  Pred. No. 0.26;
  Matches  109;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy       1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
             ||| |||| ||   ||| |  | |||||||  |||  || | ||     ||    | |
Db       2919 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 2978

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
              |  | || ||     ||    |  |  |||    ||| |||   |||| |  ||||
Db       2979 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 3038

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
                | ||||  | | ||| |  |  |  | |  |||  |    ||| |  | |||  ||
Db       3039 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 3098

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
                ||  |    ||  | || | |||  ||| |     |||
Db       3099 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 3138


RESULT 18
US-09-084-517-27
; Sequence 27, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3891 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3888
US-09-084-517-27

  Query Match             2.1%;  Score 42.4;  DB 3;  Length 3891;
  Best Local Similarity   49.5%;  Pred. No. 0.26;
  Matches  109;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy       1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
             ||| |||| ||   ||| |  | |||||||  |||  || | ||     ||    | |
Db       2919 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 2978

Qy       1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
              |  | || ||     ||    |  |  |||    ||| |||   |||| |  ||||
Db       2979 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 3038

Qy       1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
                | ||||  | | ||| |  |  |  | |  |||  |    ||| |  | |||  ||
Db       3039 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 3098

Qy       1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
                ||  |    ||  | || | |||  ||| |     |||
Db       3099 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 3138
```

RESULT 19
US-09-288-326A-10
; Sequence 10, Application US/09288326A
; Patent No. 6776990
; GENERAL INFORMATION:
; APPLICANT: Steward, Lance E.
; APPLICANT: Aoki, K. Roger
; APPLICANT: Sachs, George
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Treatment of Pancreatitis
; FILE REFERENCE: 17282
; CURRENT APPLICATION NUMBER: US/09/288,326A
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 4835
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-09-288-326A-10

Query Match 2.1%; Score 42.4; DB 3; Length 4835;
Best Local Similarity 49.5%; Pred. No. 0.29;
Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
Db 3276 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 3335

Qy 1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
Db 3336 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 3395

Qy 1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
Db 3396 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 3455

Qy 1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db 3456 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 3495

RESULT 20
US-09-548-409B-10
; Sequence 10, Application US/09548409B
; Patent No. 6843998
; GENERAL INFORMATION:
; APPLICANT: Steward,Lance E.
; APPLICANT: Aoki, K. Roger
; APPLICANT: Sachs, George
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Treament of Pancreatitis
; FILE REFERENCE: 17282CIP(AP)
; CURRENT APPLICATION NUMBER: US/09/548,409B
; CURRENT FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: US 09/288,326
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 4835
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-09-548-409B-10

Query Match 2.1%; Score 42.4; DB 3; Length 4835;
Best Local Similarity 49.5%; Pred. No. 0.29;
Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 1537 ATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTT 1596
Db 3276 ATGGAAAGTATCACTTAATTATGGTGAAATAATCTGGACTTTACAGGATACTCAGGAAAT 3335

Qy 1597 ATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAA 1656
Db 3336 AAAACAAAGAGTAGTTTTTAAATACAGTCAAATGATTAATATATCAGATTATATAAACAG 3395

Qy 1657 GCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGT 1716
Db 3396 ATGGATTTTTGTAACTATCACTAATAATAGATTAAATAACTCTAAAATTTATATAAATGG 3455

Qy 1717 TTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAAT 1756
Db 3456 AAGATTAATAGATCAAAAACCAATTTCAAATTTAGGTAAT 3495

RESULT 21
US-09-949-016-203911
; Sequence 203911, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 203911
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-203911

Query Match 2.1%; Score 42.2; DB 3; Length 601;
Best Local Similarity 49.8%; Pred. No. 0.11;
Matches 159; Conservative 0; Mismatches 158; Indels 2; Gaps 2;

Qy 1473 TGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTC 1532
Db 9 TTGTTTTTAAACAAAAATAGATTTGAAAAGAATGAACCACAACAGTGCTACCAAAAAGTC 68

Qy 1533 TCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTC 1592
Db 69 ACATCTTGTTACCAGACATCA-CTGTCCTTACAACAATTCAACTAATAAGATTTAAGATT 127

Qy 1593 ACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGA 1652
Db 128 TAACTTCATCCTAATGTATCAGAACATTAAAAAAAAAAAAGTCTCAATGTACCTATTTAA 187

Qy 1653 ACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAA-AGTCAAAATA 1711
Db 188 ATAAATCGTGCTTCCATAATAATATTCTTTAAAAATGAAAAACAGTAAAGAGCACACATT 247

Qy 1712 TTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAA 1771
Db 248 TTTATTTACTCACAACACTGAATAATTAATGTAAACTTTTTGAATTTTTTTTTCTTTAG 307

Qy 1772 CTATCTTGACTCAAGTGGA 1790
Db 308 ACATTTTTCCTCTAGAGTA 326

RESULT 22
US-09-949-016-203451
; Sequence 203451, Application US/09949016
; Patent No. 6812339

```
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 203451
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-203451

  Query Match             2.1%;  Score 42;  DB 3;  Length 601;
  Best Local Similarity   48.3%;  Pred. No. 0.13;
  Matches  114;  Conservative    1;  Mismatches  121;  Indels    0;  Gaps    0;

Qy        672 CAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCAGGCTCTTCCTC 731
              |||    ||||| |||    | |  ||| |||: | |      ||  | |  |  || ||
Db        269 CAAATCTTTCAGCTTAAAAAAAAAAAAAAAAAAMAAAAAAACAGGGCTCTGAATGATCTTC 328

Qy        732 CTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAGCATTCTTGGCA 791
                |  |||||  |   ||||| || ||| |  |||   ||       |  ||  ||| | .
Db        329 TCCTCACTGGCTCGGCTTCCCTGGCCTCATTTCCTCTATTCTCCCTACACACCCTTTGAC 388

Qy        792 AAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTGACATGTTGTAA 851
              ||  |  |  | ||      ||    || | ||  | || || ||| |    ||
Db        389 CCAGCCACACTGAATTGCTTGCAGCTCCTCAAACACGTAATCCTGTTTAATGCCTTTATG 448

Qy        852 GATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACACGAACTAA 907
              | |  |  |||||   |     || || ||| ||  |   ||  ||||   ||| |
Db        449 GTTTCAAACACATCAGGGCTAGAATGCTTTCTCCCTTTCTAAATAACATCTACTCA 504


RESULT 23
US-09-949-016-203452
; Sequence 203452, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 203452
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-203452

  Query Match             2.1%;  Score 42;  DB 3;  Length 601;
  Best Local Similarity   48.3%;  Pred. No. 0.13;
  Matches  114;  Conservative    1;  Mismatches  121;  Indels    0;  Gaps    0;

Qy        672 CAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCAGGCTCTTCCTC 731
              |||    ||||| |||    | | :||| |||  | |      ||  | |  |  || ||

Db        277 CAAATCTTTCAGCTTAAAAAAAAAAMAAAAAAAAAAAAAAAACAGGGCTCTGAATGATCTTC 336

Qy        732 CTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAGCATTCTTGGCA 791
                |  |||||  |   ||||| || ||| |  |||   ||       |  ||  ||| | .
Db        337 TCCTCACTGGCTCGGCTTCCCTGGCCTCATTTCCTCTATTCTCCCTACACACCCTTTGAC 396

Qy        792 AAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTGACATGTTGTAA 851
              ||  |  |  | ||      ||    || | ||  | || || ||| |    ||
Db        397 CCAGCCACACTGAATTGCTTGCAGCTCCTCAAACACGTAATCCTGTTTAATGCCTTTATG 456

Qy        852 GATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACACGAACTAA 907
              | |  |  |||||   |     || || ||| ||  |   ||  ||||   ||| |
Db        457 GTTTCAAACACATCAGGGCTAGAATGCTTTCTCCCTTTCTAAATAACATCTACTCA 512


RESULT 24
US-09-949-016-14773
; Sequence 14773, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14773
;   LENGTH: 113060
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-14773

  Query Match             2.1%;  Score 42;  DB 3;  Length 113060;
  Best Local Similarity   56.5%;  Pred. No. 2;
  Matches   78;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          8 TTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCAGGAC 67
              ||| ||| ||   ||  || |  ||      |  | ||| |          |||       |
Db     111243 TTTATATTAGCTCAATTAATAAGGAAATTTATTTTTATCACTACTGTATTAGCAGGATGC 111302

Qy         68 CATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATAT 127
               ||   |||||||||||||||| || | | ||||   | |||    |||||  ||| | |
Db     111303 TATCTTTTGCAAGTGACAGAAATTCTAATTAAACGGGCTTAAACGTAAAGGTGAATTTGT 111362

Qy        128 TGGCTCATGTAACCTTCT 145
              |||||| |||||   ||
Db     111363 AGGCTCAAGTAACTGACT 111380


RESULT 25
US-09-949-016-14774
; Sequence 14774, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
```

```
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14774
;   LENGTH: 113060
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-14774

  Query Match             2.1%;  Score 42;  DB 3;  Length 113060;
  Best Local Similarity   56.5%;  Pred. No. 2;
  Matches   78;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          8 TTTGTATGAGAAAAAGAAAAAGAGAGTGAAAAATATATTAACGTGCATATAGTTCAGGAC 67
              ||| ||| ||   ||  || |  ||      |  | ||| |          |||       |
Db     111243 TTTATATTAGCTCAATTAATAAGGAAATTTATTTTTATCACTACTGTATTAGCAGGATGC 111302

Qy         68 CATGGATTGCAAGTGACAGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATAT 127
              ||   |||||||||||||||| || | | ||||   | |||    |||||  ||| | |
Db     111303 TATCTTTTGCAAGTGACAGAAATTCTAATTAAACGGGCTTAAACGTAAAGGTGAATTTGT 111362

Qy        128 TGGCTCATGTAACCTTCT 145
              ||||||  |||||   ||
Db     111363 AGGCTCAAGTAACTGACT 111380


RESULT 26
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             2.1%;  Score 41.6;  DB 3;  Length 832;
  Best Local Similarity   13.2%;  Pred. No. 0.2;
  Matches   51;  Conservative  178;  Mismatches  152;  Indels    5;  Gaps    2;

Qy       1344 ATTTCTACAATTTTCCCATAAGTCATCTACACACAATACCCTCTCATGCAACACTTGGCT 1403
              | || ||| ||        ||  |  | | :: :   :::  |:|     ::      : ::
Db        390 AATTGTACCATAAATAATTATATTTTGTYYWTWWKTWWYWTTYTWTRMWWMKKKARRWYY 331

Qy       1404 TTGCTAATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCAC 1463
              ::::| | |:::::::|: ::: :  :: :|: : :  :  | :: :   ::|: : :::
Db        330 WWKSTYACASRYRKYTWGWWWYMWKRMMSTRWYCYMCWKCCMYRGRRCAWYTMARGRMWS 271

Qy       1464 TAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCT 1523
              :|:|:    ::::  |:   |::  ::|:| :|     ||   :::    : :::| ::
Db        270 YAWGK----WKSMRSAMSMCTRMYYKKGSTYWTMKCTCATWCYWYWKYWKRMWSKTCWSG 215

Qy       1524 CCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATT 1583
              :          :|:   ::  :|::::|: :::|:::: :: :|:  ::    || ::  ::::
Db        214 SRGGYMTSYTSTRSYSMYWASWMYTMCWWWGRWWSTYWYMAWGKKWWRYATTWRRAMMWW 155

Qy       1584 TCTCTGCTCACTTATCATA-AGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGT 1642
              :   | :::: : :  : |   ::|:: :   :::::  ::: :  :|:|:: ::  :
Db        154 WAAWTMMWYMWWAWCMSSRGAAMYRRTMMMWGYRYWWRKKSYRRTRCAWAYAWKTKRSYY 95
```

```
Qy       1643 AATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAA 1702
              : : ::: : :::: :   : :| :  | : |: ::::|: :::::  :::: : : :|:
Db         94 WCWRWKWKRCMMMMMMAMAYGKTMMMRACWKTRYWRWWAWAMWRMWWTMMMMYYWYWRAM 35

Qy       1703 GTCAAAATATTTGTTTAATCAGGTCA 1728
              :  ::::::  ::: ::|: : :: |
Db         34 KRRWMWRKWRSWSWMWMAWGMTRWAA 9


RESULT 27
US-08-049-473-1/c
; Sequence 1, Application US/08049473
; Patent No. 5386021
;  GENERAL INFORMATION:
;    APPLICANT:  Moss, Joel
;    APPLICANT:  Mishima, Koichi
;    APPLICANT:  Nightingale, Maria
;    APPLICANT:  Tsuchiya, Mikako
;    TITLE OF INVENTION:  A MAMMALIAN GUANIN NUCLEOTIDE BINDING
;    TITLE OF INVENTION:  PROTEIN WITH AN ADP-RYBOSYLATION FACTOR DOMAIN
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KNOBBE, MARTENS, OLSON AND BEAR
;      STREET:  620 NEWPORT CENTER DRIVE  SIXTEENTH FLOOR
;      CITY:  NEWPORT BEACH
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,473
;      FILING DATE:  19930419
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fuller, Michael L.
;      REGISTRATION NUMBER:  36,516
;      REFERENCE/DOCKET NUMBER:  NIH050.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3312 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1725
US-08-049-473-1

  Query Match             2.1%;  Score 41.2;  DB 2;  Length 3312;
  Best Local Similarity   48.8%;  Pred. No. 0.53;
  Matches  169;  Conservative    0;  Mismatches  173;  Indels    4;  Gaps    2;

Qy       1446 TAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGT 1505
              | |||| ||| | | |   ||       | | ||  | | ||||   |  ||||  |   |
Db       3118 TTAATGCTTTGTTTTCTAAAACTTGTATTGTTTTTAAACAAAAATAGATTTGAAAAGAAT 3059

Qy       1506 GATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCT-GAA 1564
              ||  || ||   |  |  | | | ||| ||| |   |  |  ||| |   |   | ||
Db       3058 GAACCACAACAGTGCTACCAAAAAGTCACATCTTGTTACCAGACATCACTGTCCTTACAA 2999

Qy       1565 ATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAAC 1624
              |  |   |       ||||       | ||||       |      |||  |  | |  ||
Db       2998 CAATTCAACTAATAAGATTTAAGATTTAACTTCATCCTAATGTATCAGAACATTAAAAAA 2939
```

```
Qy       1625 AAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGA 1684
             ||||    |||| ||  || ||   | |  || |  | ||   | |  |||    | | |
Db       2938 AAAAGTCTCAATGTACCTATTT---AAATAAATCGTGCTTCCATAATAATATTCTTTAAA 2882

Qy       1685 AAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAA 1744
             ||   ||| ||    | ||   | || ||| |||| |||   || |||  |  || |  |
Db       2881 AATGAAAAACAGTAAAGAGCACACATTTTTATTTACTCACAACACTGAATAATTAATGTA 2822

Qy       1745 AAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGA 1790
             || | ||||||||| ||| |       |   || ||  ||| || | |
Db       2821 AACTTTTTGAATTTTTTTTTTTCTTTAGACATTTTTCCTCTAGAGTA 2776


RESULT 28
US-08-312-648-1/c
; Sequence 1, Application US/08312648
; Patent No. 5514600
;  GENERAL INFORMATION:
;    APPLICANT:  Moss, Joel
;    APPLICANT:  Mishima, Koichi
;    APPLICANT:  Nightingale, Maria
;    APPLICANT:  Tsuchiya, Mikako
;    TITLE OF INVENTION:  A MAMMALIAN GUANIN NUCLEOTIDE BINDING
;    TITLE OF INVENTION:  PROTEIN WITH AN ADP-RYBOSYLATION FACTOR DOMAIN
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KNOBBE, MARTENS, OLSON AND BEAR
;      STREET:  620 NEWPORT CENTER DRIVE  SIXTEENTH FLOOR
;      CITY:  NEWPORT BEACH
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/312,648
;      FILING DATE:
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/049,473
;      FILING DATE:  19-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fuller, Michael L.
;      REGISTRATION NUMBER:  36,516
;      REFERENCE/DOCKET NUMBER:  NIH050.001DV1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3312 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1725
US-08-312-648-1

  Query Match             2.1%;  Score 41.2;  DB 2;  Length 3312;
  Best Local Similarity   48.8%;  Pred. No. 0.53;
  Matches  169;  Conservative    0;  Mismatches  173;  Indels    4;  Gaps    2;

Qy       1446 TAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGT 1505
             | |||| ||| | | |   ||     | | ||  | | ||||   |  ||||  |   |
Db       3118 TTAATGCTTTGTTTTCTAAAACTTGTATTGTTTTTAAACAAAAATAGATTTGAAAAGAAT 3059

Qy       1506 GATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCT-GAA 1564
             ||  || ||   | | | | | ||| ||| |    |  | ||| |   |   |  ||
Db       3058 GAACCACAACAGTGCTACCAAAAAGTCACATCTTGTTACCAGACATCACTGTCCTTACAA 2999

Qy       1565 ATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAAC 1624
               | |   |       ||||      | ||||      |     |||  |  | |  ||
Db       2998 CAATTCAACTAATAAGATTTAAGATTTAACTTCATCCTAATGTATCAGAACATTAAAAAA 2939

Qy       1625 AAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGA 1684
             ||||    |||| ||  || ||   | |  || |  | ||   | |  |||    | | |
Db       2938 AAAAGTCTCAATGTACCTATTT---AAATAAATCGTGCTTCCATAATAATATTCTTTAAA 2882

Qy       1685 AAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAA 1744
             ||   ||| ||    | ||   | || ||| |||| |||   || |||  |  || |  |
Db       2881 AATGAAAAACAGTAAAGAGCACACATTTTTATTTACTCACAACACTGAATAATTAATGTA 2822

Qy       1745 AAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGA 1790
             || | ||||||||| ||| |       |   || ||  ||| || | |
Db       2821 AACTTTTTGAATTTTTTTTTTTCTTTAGACATTTTTCCTCTAGAGTA 2776


RESULT 29
PCT-US94-04190-1/c
; Sequence 1, Application PC/TUS9404190
;  GENERAL INFORMATION:
;    APPLICANT:  The Government of the United States of America
;    APPLICANT:  as represented by the Secretary, Department
;    APPLICANT:  of Health and Human Services
;    TITLE OF INVENTION:  A MAMMALIAN GUANIN NUCLEOTIDE BINDING
;    TITLE OF INVENTION:  PROTEIN WITH AN ADP-RYBOSYLATION FACTOR DOMAIN
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KNOBBE, MARTENS, OLSON AND BEAR
;      STREET:  620 NEWPORT CENTER DRIVE  SIXTEENTH FLOOR
;      CITY:  NEWPORT BEACH
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04190
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fuller, Michael L.
;      REGISTRATION NUMBER:  36,516
;      REFERENCE/DOCKET NUMBER:  NIH050.001QPC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3312 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;    FEATURE:  7          (A) NAME/KEY: CDS
;      LOCATION:  1..1725
PCT-US94-04190-1

  Query Match             2.1%;  Score 41.2;  DB 6;  Length 3312;
  Best Local Similarity   48.8%;  Pred. No. 0.53;
  Matches  169;  Conservative    0;  Mismatches  173;  Indels    4;  Gaps    2;
```

```
Qy       1446 TAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGT 1505
             | |||| ||| | | |   ||      | | ||  | | ||||   |  ||||  |    |
Db       3118 TTAATGCTTTGTTTTCTAAAACTTGTATTGTTTTTAAACAAAAATAGATTTGAAAAGAAT 3059

Qy       1506 GATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGCT-GAA 1564
             ||  || ||   |  |  | | | ||| ||| |    |  | ||| |   |   |  ||
Db       3058 GAACCACAACAGTGCTACCAAAAAGTCACATCTTGTTACCAGACATCACTGTCCTTACAA 2999

Qy       1565 ATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAAC 1624
               |  |   |      ||||      | ||||       |     |||  |  | |  ||
Db       2998 CAATTCAACTAATAAGATTTAAGATTTAACTTCATCCTAATGTATCAGAACATTAAAAAA 2939

Qy       1625 AAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGA 1684
             ||||   |||| ||  || ||   | |  || |  | ||   | |  |||    | | |
Db       2938 AAAAGTCTCAATGTACCTATTT---AAATAAATCGTGCTTCCATAATAATATTCTTTAAA 2882

Qy       1685 AAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAAA 1744
             ||   ||| ||    | |||   | || ||| |||| |||   || |||   |  || |  |
Db       2881 AATGAAAAACAGTAAAGAGCACACATTTTTATTTACTCACAACACTGAATAATTAATGTA 2822

Qy       1745 AAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGA 1790
             || | |||||||| ||| |       |   || ||  ||| || | |
Db       2821 AACTTTTTGAATTTTTTTTTTCTTTAGACATTTTTCCTCTAGAGTA 2776
```

RESULT 30
US-09-949-016-14773/c
; Sequence 14773, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14773
; LENGTH: 113060
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14773

Query Match 2.1%; Score 41.2; DB 3; Length 113060;
Best Local Similarity 43.8%; Pred. No. 3.3;
Matches 178; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

```
Qy       1441 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 1500
             || |  |||| | |||  |  || |   |||    | || || ||||| | || ||
Db     106708 AACCACAAATTTATTAACTTTACCATCCCTCCAACCATATATGTAAAAAGAGTCTTTTAG 106649

Qy       1501 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 1560
             |||  |    |  |  || ||||  || | |   ||     | |  |  | |  ||
Db     106648 AATTTCTAGCCCATATCATTTCCTCATATGATGCAAGACCATAACACCCTTCTTATACTA 106589

Qy       1561 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 1620
             | | || ||  || |      |||  |    |       | |         | |      ||
Db     106588 TTATATGGATTGCTTCACGTATTAATACCTTTCTCCTTTACCCATGTCTAACCTAAGAAA 106529

Qy       1621 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 1680
               | |   ||    | | ||  || | ||||||       || ||  |||  ||| ||
Db     106528 TCAGAGGGGCTTTTCTTTTTGGTATAACAAGAACTCTAAAAATTAGCAAAATGAAAATAT 106469

Qy       1681 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 1740
               |      |  |  | | |||||   || ||||  |   | |  ||   | ||    ||
Db     106468 CTTATTTGAATCCTTTCTAAAAGTTGGAACATTTAGTCTCTTATTTCTAAGGGACCAAAT 106409

Qy       1741 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 1800
             ||||     | |  || ||||  | ||     |   |       | |  |   ||  | ||
Db     106408 TAAATTTAAATGAAACTCTTCTTTATATTCATGGTAACAGGTTACTTATCCAAAGAAGAT 106349

Qy       1801 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAA 1846
               |      |   ||||| |  | || |  |  |   | || ||
Db     106348 TTAGGCTGGAAGTGGTGCACATGTGGGCAATATATTCATTTATGAA 106303
```

RESULT 31
US-09-949-016-14774/c
; Sequence 14774, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14774
; LENGTH: 113060
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14774

Query Match 2.1%; Score 41.2; DB 3; Length 113060;
Best Local Similarity 43.8%; Pred. No. 3.3;
Matches 178; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

```
Qy       1441 AAGCGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGC 1500
             || |  |||| | |||  |  || |   |||    | || || ||||| | || ||
Db     106708 AACCACAAATTTATTAACTTTACCATCCCTCCAACCATATATGTAAAAAGAGTCTTTTAG 106649

Qy       1501 AATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGATTGGC 1560
             |||  |    |  |  || ||||  || | |   ||     | |  |  | |  ||
Db     106648 AATTTCTAGCCCATATCATTTCCTCATATGATGCAAGACCATAACACCCTTCTTATACTA 106589

Qy       1561 TGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAG 1620
             | | || ||  || |      |||  |    |       | |         | |      ||
Db     106588 TTATATGGATTGCTTCACGTATTAATACCTTTCTCCTTTACCCATGTCTAACCTAAGAAA 106529

Qy       1621 GAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAAT 1680
               | |   ||    | | ||  || | ||||||       || ||  |||  ||| ||
Db     106528 TCAGAGGGGCTTTTCTTTTTGGTATAACAAGAACTCTAAAAATTAGCAAAATGAAAATAT 106469

Qy       1681 AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTAT 1740
               |      |  |  | | |||||   || ||||  |   | |  ||   | ||    ||
Db     106468 CTTATTTGAATCCTTTCTAAAAGTTGGAACATTTAGTCTCTTATTTCTAAGGGACCAAAT 106409

Qy       1741 TAAAAAGTATTTGAATTCTTTATGATGAGAACTATCTTGACTCAAGTGGACAGTGGTGAG 1800
             ||||     | |  || ||||  | ||     |   |       | |  |   ||  | ||
Db     106408 TAAATTTAAATGAAACTCTTCTTTATATTCATGGTAACAGGTTACTTATCCAAAGAAGAT 106349

Qy       1801 CTTTTTGGCCTGTGGTCCCTACGTAGAAAGGAGGCTTTGTCATAAA 1846
               |      |   ||||| |  | || |  |  |   | || ||
Db     106348 TTAGGCTGGAAGTGGTGCACATGTGGGCAATATATTCATTTATGAA 106303
```

RESULT 32
US-09-949-016-14043/c
; Sequence 14043, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

```
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14043
;   LENGTH: 312470
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(312470)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14043

  Query Match             2.0%;  Score 41;  DB 3;  Length 312470;
  Best Local Similarity   47.8%;  Pred. No. 6.5;
  Matches  119;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy       1536 TATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACT 1595
             ||| |||    | |     |   || |||||   | |   ||| | || ||||     |
Db      19287 TATCAAAATTCATAAAACTGTCCACTTAAATAAGTTAAGAAGTCAAATTTTCTGTATTAT 19228

Qy       1596 TATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACA 1655
              ||  | |  ||   | |    |||  ||||||| || ||  | |  | | | || |||
Db      19227 AATTTTTAAAAATAGAATTAAGAAGTCACAAAAATTAAAAAATGTTAACTAACAAAAACC 19168

Qy       1656 AGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTG 1715
              |  |   | |       |||||    | |||   | | |   | |     | |    | ||
Db      19167 AATTTAATCTTGCAGTAATAAACTGGACAAACCCATAGCTATTTATGACTAGATGGATGG 19108

Qy       1716 TTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGAGAACTAT 1775
              |||   |       | |||| || |  ||| ||||| |  |   |||||     |  ||
Db      19107 TTTCCACTATCTTCAGGGAATTTACTTTAAAATATTTTATATACGTATGAACCAGATCAT 19048

Qy       1776 CTTGACTCA 1784
              ||| || |
Db      19047 GTTGTCTAA 19039


RESULT 33
US-09-949-016-17484/c
; Sequence 17484, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17484
;   LENGTH: 92155
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17484

  Query Match             2.0%;  Score 40.8;  DB 3;  Length 92155;
  Best Local Similarity   48.3%;  Pred. No. 3.9;
  Matches  114;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy        672 CAATGTCTTCAGATTATCCTAGACCAAATAAACTACAGCCACTGGGCCAGGCTCTTCCTC 731
              |||    ||||| |||    | |  ||| |||  | |       ||  | |  |  || ||
Db      23679 CAAATCTTTCAGCTTAAAAAAAAAAAAAAAAAAAAAAAAAAACAGGGCTCTGAATGATCTTC 23620

Qy        732 CTCCAACTGGACCATGTTCCCAGGGCTCTTCACCTTAGTTTAGGTCAAGCATTCTTGGCA 791
               |  |||||  |   ||||| || ||| |  |||   ||       |  ||  ||| | |
Db      23619 TCCTCACTGGCTCGGCTTCCCTGGCCTCATTTCCTCTATTCTCCCTACACACCCTTTGAC 23560

Qy        792 AAAGAAAGGCCTAGTTAACAATAGACATTCTAGCAATTGATTCTTTTTGACATGTTGTAA 851
              ||  |  |  | ||     ||    || | ||  | || || ||| |    ||
Db      23559 CCAGCCACACTGAATTGCTTGCAGCTCCTCAAACACGTAATCCTGTTTAATGCCTTTATG 23500

Qy        852 GATCTATTCACATTTTGTAATTAAAGCATTCCCCTATGGAAACCAACACGAACTAA 907
              | |  |  |||||   |     || || ||| ||  |   ||  ||||   ||| |
Db      23499 GTTTCAAACACATCAGGGCTAGAATGCTTTCTCCCTTTCTAAATAACATCTACTCA 23444


RESULT 34
US-09-949-016-86304/c
; Sequence 86304, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 86304
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-86304

  Query Match             2.0%;  Score 40.6;  DB 3;  Length 601;
  Best Local Similarity   50.3%;  Pred. No. 0.32;
  Matches  100;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy       1564 AATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAA 1623
              |||  | | || | |  |||| |   |  || |||   ||  |  || ||| |    ||
Db        515 AATTTATTCATTTATTTATTTTTGAACATACATATTCAAAAAATATCTTTATATTTTTAA 456

Qy       1624 CAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAG 1683
               | | |  ||| | |     |||| || ||  |       |  || ||||||  || |
Db        455 TATAGAAAACATTTTTGTATATTAAAATAAATATAAAACATATCTTAATATATTAAAATA 396

Qy       1684 AAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAA 1743
              |||| |    || | ||  |  | | |||   |  | |  |  || |  || ||  |
Db        395 TAAAACATTATATATAAATATATAGAAATTATATATTTTATATATTTAAATATATAAAAC 336

Qy       1744 AAAGTATTTGAATTCTTTA 1762
              |   ||||| |  | ||||
Db        335 AGTATATTTTATATATTTA 317


RESULT 35
US-09-949-016-86305/c
; Sequence 86305, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
```

```
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 86305
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-86305

  Query Match             2.0%;  Score 40.6;  DB 3;  Length 601;
  Best Local Similarity   50.3%;  Pred. No. 0.32;
  Matches  100;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy       1564 AATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAA 1623
             ||| | | || | |  |||| |    |  || |||   ||  |  || |||  |    ||
Db        546 AATTTATTCATTTATTTATTTTTGAACATACATATTCAAAAAATATCTTTATATTTTTAA 487

Qy       1624 CAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAG 1683
              | | |  ||| | |     |||| || ||  |       |  || ||||||  || |
Db        486 TATAGAAAACATTTTTGTATATTAAAATAAATATAAAACATATCTTAATATATTAAAATA 427

Qy       1684 AAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAA 1743
              |||| |    || | ||  |  | | |||   |   | |  |  || |  || ||  |
Db        426 TAAAACATTATATATAAATATATAGAAATTATATATTTTATATATTTAAATATATAAAAC 367

Qy       1744 AAAGTATTTGAATTCTTTA 1762
              |   ||||| |  | ||||
Db        366 AGTATATTTTATATATTTA 348


RESULT 36
US-09-949-016-14182
; Sequence 14182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14182
;   LENGTH: 422592
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(422592)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14182

  Query Match             2.0%;  Score 40.6;  DB 3;  Length 422592;
  Best Local Similarity   50.3%;  Pred. No. 9.9;
  Matches  100;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy       1564 AATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAGCTAAGGAA 1623
             ||| | | || | |  |||| |    |  || |||   ||  |  || |||  |    ||
Db     333066 AATTTATTCATTTATTTATTTTGAACATACATATTCAAAAAATATCTTTATATTTTTAA 333125

Qy       1624 CAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAG 1683
              | | |  ||| | |     |||| || ||  |       |  || ||||||  || |
Db     333126 TATAGAAAACATTTTTGTATATTAAAATAAATATAAAACATATCTTAATATATTAAAATA 333185

Qy       1684 AAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAATCTATTAA 1743
              |||| |    || | ||  |  | | |||   |   | |  |  || |  || ||  |
Db     333186 TAAAACATTATATATAAATATATAGAAATTATATATTTTATATATTTAAATATATAAAAC 333245

Qy       1744 AAAGTATTTGAATTCTTTA 1762
              |   ||||| |  | ||||
Db     333246 AGTATATTTTATATATTTA 333264


RESULT 37
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             2.0%;  Score 40.6;  DB 3;  Length 640681;
  Best Local Similarity   48.8%;  Pred. No. 12;
  Matches  139;  Conservative    0;  Mismatches  144;  Indels    2;  Gaps    1;

Qy       1405 TGCTAATACATATCTATTATGAGAGCTGTGCTTCTTAAGCGTAAATGTTTTATATGCACT 1464
              | ||  |  | | |    || | ||   || |  ||||   |     |||||| ||  ||
Db      65132 TTCTGTTTTAAACCATCAATAAAAGTACTGTTATTTAATTTTTGTATTTTTATTTGGTCT 65073

Qy       1465 AAGGCTCTTGGCTTACATATAAAAGGGGTATTGAGCAATGTGATACAGAAGTCTTTTCTC 1524
              |    |    |    | | |||||          || ||  |  || |  | | | || |
Db      65072 ATTTTTTCAAGTAAGCTTTTAAAAAAAAATCGAAGTAAAATTTTATATCAATATGTTGTT 65013

Qy       1525 CACAGGTCTCATATGTAAAGAATTCATTAGATTGGCTGAAATAGACTGATCTGTCCATTT 1584
              | |   || |||  |||| |||| |  | ||    |  |||| | | || | | ||||
Db      65012 TAGAATGCTAATAGATAAAAAATTGAAAAAATATTATTCAATATAATAATTTCTACATTG 64953

Qy       1585 CTCTGCTCACTTATC--ATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGT 1642
              ||||       | ||   |||   ||   |||| |||   ||  | ||      | |||  |
Db      64952 CTCTTAAATTTAATTCAATATATAATATATTAACTATACAAACATAAAATTTAACTAATT 64893

Qy       1643 AATTAGAAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAA 1687
              |    | |  || ||  ||||    ||| |||   ||  || |
Db      64892 TTTCTATAAATAAAACTTTTTTTAAATAATTTAACTCTATTAAGA 64848


RESULT 38
US-09-248-796A-10353/c
; Sequence 10353, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
```

```
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 10353
;   LENGTH: 252
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-09-248-796A-10353

  Query Match             2.0%;  Score 40.4;  DB 3;  Length 252;
  Best Local Similarity   53.1%;  Pred. No. 0.23;
  Matches   86;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       1602 AAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGG 1661
              || |   | || || |||  || |  || | |||   |        ||| | | | |||
Db        228 AATGTCATAATGAGATAATAAATAGGAATTCCAAAAAAATAGGGATGAAAAGCTAACTGA 169

Qy       1662 TTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAAT 1721
              |||   | ||  ||| || ||||  ||| || | ||  |||   |||| ||      |||
Db        168 TTTATATGAAAGTAATAACAAGAGGAAGGAATCTTGAAAAATGAAAAAAATAAAAAAAAT 109

Qy       1722 CAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTAT 1763
              || |     | ||  | | |   | |||  | ||| ||| |
Db        108 TAGATGTGGTACAAAATTTAATTGATTATGAGTATTTTTTTT 67


RESULT 39
US-09-573-080A-118/c
; Sequence 118, Application US/09573080A
; Patent No. 6828097
; GENERAL INFORMATION:
;  APPLICANT:  JOAN, KNOLL
;  APPLICANT:  ROGAN, PETER
;  TITLE OF INVENTION:   SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATI
;  FILE REFERENCE:   30307
;  CURRENT APPLICATION NUMBER: US/09/573,080A
;  CURRENT FILING DATE:  2000-05-16
;  NUMBER OF SEQ ID NOS:   479
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 118
;   LENGTH: 476
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(476)
;  PUBLICATION INFORMATION:
;  PUBLICATION INFORMATION:
;   AUTHORS: Jurka, J; Walichiewicz, J; Milosavljevic, A
;   TITLE: Prototypic sequences for human repetitive DNA
;   JOURNAL: Journal of Molecular Evolution
;   VOLUME: 35
;   ISSUE: 4
;   PAGES: 286-291
;   DATE: 1992-10-__
;   DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
;   DATABASE ENTRY DATE: ____-__-__
;   DATABASE ENTRY DATE: 1996-01-26
US-09-573-080A-118

  Query Match             2.0%;  Score 39.6;  DB 3;  Length 476;
  Best Local Similarity   80.4%;  Pred. No. 0.54;
  Matches   45;  Conservative    1;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         85 AGAAACTCAATTCAAACCAACGTAAGTCAAAAGGAAAATATATTGGCTCATGTAAC 140
              |||||| ||| |||||| : | ||||  |||| || ||| ||||||||||||||||
Db        461 AGAAACCCAACTCAAACTRGCTTAAGCAAAAAAGAGAATTTATTGGCTCATGTAAC 406


RESULT 40
US-09-949-002-645
; Sequence 645, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 645
;   LENGTH: 56374
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-002-645

  Query Match             2.0%;  Score 39.6;  DB 3;  Length 56374;
  Best Local Similarity   47.6%;  Pred. No. 6.7;
  Matches  117;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy       1496 TGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGA 1555
              | |  ||||| ||          ||  |  ||| || | ||||   |  |||| | |
Db      29627 TTAAAAATGTTATTTCTCTTAGATTAATATACAAGTTTAATATAATACCAATTAAATTAC 29686

Qy       1556 TTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAG 1615
                     | |||  ||    ||| |||    |  ||    | || |  |  |  ||
Db      29687 CAAATATCCTAAAACTTGTCAAAACATGTCTAAAATTTCTGTCAAAAATGCTGGAAACAG 29746

Qy       1616 CTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATA 1675
              ||   || ||||| |  ||| | | ||  |||    |    || ||     | || |||
Db      29747 TTAGCCAAGAAAAAATTGAATAGAAGAAAAAAGAGCCATTTCCTTGTAAGTATTAAAATA 29806

Qy       1676 AAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAA 1735
              |    | | | ||| |  ||  | | ||||| | | ||| | | |  || |    || |
Db      29807 TATTATAAAGATAGAGATTATTAGTATGTCAACAAAATTGCTGACTAGGGGCTCCTAGCA 29866

Qy       1736 TCTATT 1741
                 |||
Db      29867 CTAATT 29872


RESULT 41
US-09-949-002-774
; Sequence 774, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 774
;   LENGTH: 56375
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-002-774

  Query Match             2.0%;  Score 39.6;  DB 3;  Length 56375;
  Best Local Similarity   47.6%;  Pred. No. 6.7;
```

```
  Matches  117;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy       1496 TGAGCAATGTGATACAGAAGTCTTTTCTCCACAGGTCTCATATGTAAAGAATTCATTAGA 1555
             | |  ||||| ||          ||   |  ||| || | ||||    |  |||| | |
Db      29627 TTAAAAATGTTATTTCTCTTAGATTAATATACAAGTTTAATATAATACCAATTAAATTAC 29686

Qy       1556 TTGGCTGAAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAAGGAAGTCATTAG 1615
                     | |||  ||    ||| |||     |  ||    | || |  |  |  ||
Db      29687 CAAATATCCTAAAACTTGTCAAAACATGTCTAAAATTTCTGTCAAAAATGCTGGAAACAG 29746

Qy       1616 CTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTGCTCAATATA 1675
               ||   || ||||| |  |||  | | ||  |||    |     || ||     | || |||
Db      29747 TTAGCCAAGAAAAAATTGAATAGAAGAAAAAGAGCCATTTCCTTGTAAGTATTAAAATA 29806

Qy       1676 AAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGTCATTGAGAA 1735
               |    | | | ||| |  ||  | | ||||| | | ||| | | |  || |    || |
Db      29807 TATTATAAAGATAGAGATTATTAGTATGTCAACAAAATTGCTGACTAGGGGCTCCTAGCA 29866

Qy       1736 TCTATT 1741
                 |||
Db      29867 CTAATT 29872


RESULT 42
US-08-096-181A-7
; Sequence 7, Application US/08096181A
; Patent No. 6153406
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level Expression,
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/096,181A
;      FILING DATE:  23-Jul-1993
;      CLASSIFICATION:  424
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  65..1147
US-08-096-181A-7

  Query Match             2.0%;  Score 39.4;  DB 3;  Length 1477;
  Best Local Similarity   54.5%;  Pred. No. 1.1;
  Matches   79;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy       1607 AGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTG 1666
             | ||||| | ||    |   | | || || | | |  | | |||  |   |||  |||||
Db       1149 AATCATTTGTTAGAAATACATTATTAAAAGCAAGGCGAATCGAAAGATTCGCTTTTTTTG 1208

Qy       1667 CTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGT 1726
             ||||| || ||  ||| |||      |   |   |  ||  ||||||||      ||    |
Db       1209 CTCAAAATCAAGTTAAAAAATGATTAAGTTAAAAGTGTATAAATATTTAGGCTATTTTAT 1268

Qy       1727 CATTGAGAATCTATTAAAAAGTATT 1751
              | | | ||  |||||| ||  | |
Db       1269 AAGTAACAAAATATTAATAAAAAAT 1293


RESULT 43
PCT-US94-08326-7
; Sequence 7, Application PC/TUS9408326
;  GENERAL INFORMATION:
;    APPLICANT:  North American Vaccine, Inc.
;    APPLICANT:  12103 Indian Creek Court
;    APPLICANT:  Beltsville, MD 20705
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level
;    TITLE OF INVENTION:  Expression,
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane
;    TITLE OF INVENTION:  Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/08326
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/096,181
;      FILING DATE:  23-JULY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REFERENCE/DOCKET NUMBER:  1438.001PC01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  65..1147
PCT-US94-08326-7

  Query Match             2.0%;  Score 39.4;  DB 6;  Length 1477;
  Best Local Similarity   54.5%;  Pred. No. 1.1;
  Matches   79;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy       1607 AGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAGAAGAACAAGCTGGTTTTG 1666
             | ||||| | ||    |   | | || || | | |  | | |||  |   |||  |||||
Db       1149 AATCATTTGTTAGAAATACATTATTAAAAGCAAGGCGAATCGAAAGATTCGCTTTTTTTG 1208

Qy       1667 CTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGGT 1726
             ||||| || ||  ||| |||      |   |   |  ||  ||||||||      ||    |
Db       1209 CTCAAAATCAAGTTAAAAAATGATTAAGTTAAAAGTGTATAAATATTTAGGCTATTTTAT 1268
```

```
Qy       1727 CATTGAGAATCTATTAAAAAGTATT 1751
              | | | ||  |||||| ||  | |
Db       1269 AAGTAACAAAATATTAATAAAAAAT 1293
```

RESULT 44
US-09-350-756-8
; Sequence 8, Application US/09350756
; Patent No. 6495143
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Reseach Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 8
; LENGTH: 1987
; TYPE: DNA
; ORGANISM: Clostridium botulinum
; FEATURE:
US-09-350-756-8

Query Match 2.0%; Score 39.2; DB 3; Length 1987;
Best Local Similarity 51.1%; Pred. No. 1.5;
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy       1589 GCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAG 1648
              | ||  | |  |||  | |   | | | | || ||     | ||||     ||   | |
Db        825 GGTCTATAAATATATAGTAACAAATTGGTTAGCAAAGGTTAATACACAGATTGATCTAAT 884

Qy       1649 AAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAA 1708
              ||||| ||    |       | || ||||  ||| | ||| ||| | | |   |   |||
Db        885 AAGAAAAAAAAATGAAAGAAGCTTTAGAAAATCAAGCAGAAGCAACAAAGGCTATAATAAA 944

Qy       1709 ATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGA 1768
              |||  || ||||||    | |||| |       ||| | ||||    ||  |  ||||||
Db        945 CTATCAGTATAATCAATATACTGAGGAAGAGAAAAATAATATTAATTTTAATATTGATGA 1004
```

RESULT 45
US-09-350-756-7
; Sequence 7, Application US/09350756
; Patent No. 6495143
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Reseach Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 7
; LENGTH: 2452
; TYPE: DNA
; ORGANISM: Clostridium botulinum
; FEATURE:
US-09-350-756-7

Query Match 2.0%; Score 39.2; DB 3; Length 2452;
Best Local Similarity 51.1%; Pred. No. 1.7;
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy       1589 GCTCACTTATCATAAGGAAGTCATTAGCTAAGGAACAAAAACTACAATCTATGTAATTAG 1648
              | ||  | |  |||  | |   | | | | || ||     | ||||     ||   | |
Db        705 GGTCTATAAATATATAGTAACAAATTGGTTAGCAAAGGTTAATACACAGATTGATCTAAT 764

Qy       1649 AAGAACAAGCTGGTTTTGCTCAATATAAAAATAAGAAAAAGAAACCATGTGAAAGTCAAA 1708
              ||||| ||    |       | || ||||  ||| | ||| ||| | | |   |   |||
Db        765 AAGAAAAAAAAATGAAAGAAGCTTTAGAAAATCAAGCAGAAGCAACAAAGGCTATAATAAA 824

Qy       1709 ATATTTGTTTAATCAGGTCATTGAGAATCTATTAAAAAGTATTTGAATTCTTTATGATGA 1768
              |||  || ||||||    | |||| |       ||| | ||||    ||  |  ||||||
Db        825 CTATCAGTATAATCAATATACTGAGGAAGAGAAAAATAATATTAATTTTAATATTGATGA 884
```

Search completed: January 22, 2006, 21:30:21
Job time : 255.225 secs

This Page Blank (uspto)